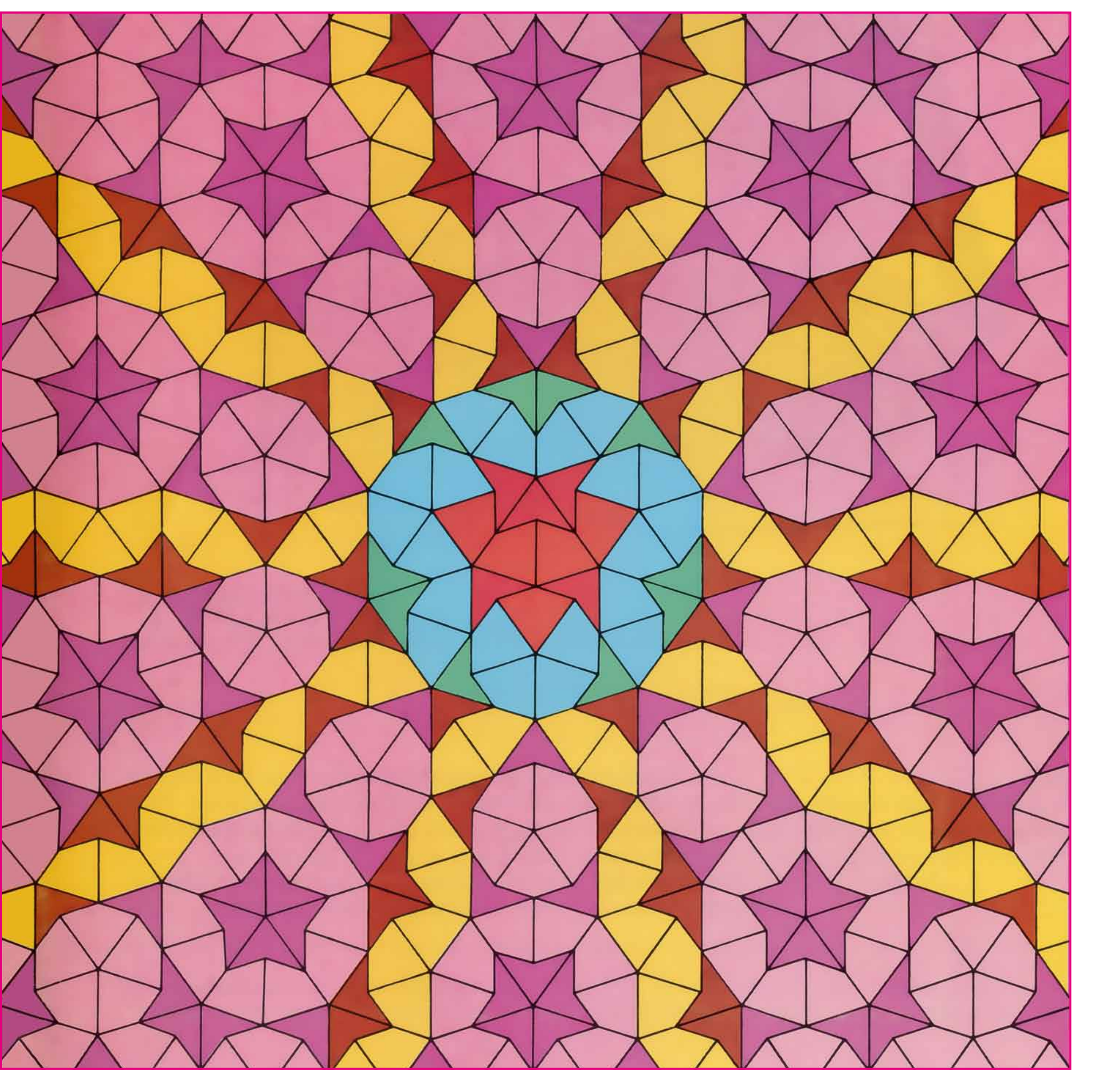
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



TEORIA DEL MOSAICO

Marzo 1977

140 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 6
Marzo 1977

INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-15
ESPAÑA

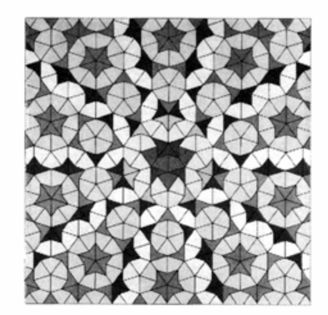
Colaboradores de este número:

Asesoramiento y traducción:

Pilar Martínez: *El aborto legal;* Francisco Gil: *Agricultura sin labranza;* Enric Verdaguer: *La mecánica cuántica de los agujeros negros;* Salvador Carmaniu: *El lado de unión de los anticuerpos;* Monique Robert: *La percepción de objetos en movimiento;* Nuria Ferrer: *Exoelectrones;* Manuel Puigcerver: *Formación de cráteres en el sistema solar;* Ramón Fontarnau y Frances E. Lynd: *Taller y laboratorio;* Enrique Trillas y Teresa Riera: *Juegos matemáticos.*

Libros:

Philip Morrison



LA PORTADA

La ilustración de la portada es un mosaico no periódico descubierto por el matemático inglés Roger Penrose (véase "Juegos matemáticos"). Mosaico periódico es aquel en que una transparencia del mosaico modelo puede colocarse sobre el mismo modelo y desplazarse a infinitas nuevas posiciones (sin rotación ni reflexión), sin dejar de encajar en el patrón. Son mosaicos periódicos típicos las baldosas hexagonales de un cuarto de baño y los cuadros de un tablero de damas. Hay formas más libres, como las que aparecen en los dibujos repetitivos del artista holandés M.C. Escher, que también pueden prestarse para mosaicos periódicos. Los mosaicos no periódicos pueden llenar también el plano con modelos que posean varios tipos de simetría, pero carecen de una "región fundamental" que embaldose por traslación. Algunos mosaicos no periódicos, como el de la portada, poseen un centro único desde el cual rasgos, tales como los "rayos de rueda", se proyectan como radios hacia el infinito.

Suscripciones:

Prensa Científica, S.A.
Departamento de suscripciones
Calabria, 235-239
Barcelona-15 (España)

Distribución para España:

Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:

Editorial Labor, S.A.
Calabria, 235-239
Barcelona-15 (España)

Condiciones de suscripción:

España:
Un año (12 números): 1.540 pesetas
Seis meses (6 números): 800 pesetas

Extranjero:
Un año (12 números); 30 U.S. $

Ejemplar atrasado ordinario:
160 pesetas
Ejemplar atrasado extraordinario:
225 pesetas.

PROCEDENCIA DE LAS ILUSTRACIONES

Diseño de la cubierta por Andrew Christie

Página	Fuente	Página	Fuente
6-13	Jerome Kuhl	58-63	Gabor Kiss
15	Allis-Chalmers	64	Ernest Rabinowicz y Gabor Kiss
16	Alan D. Iselin	65	William J. Baxter y Stanley R. Rouse
17	Ohio Agricultural Research and Development Center	67	NASA y U.S. Geological Survey
18-20	Alan D. Iselin	68	NASA y Hale Observatories
23-28	Dan Todd	69	NASA
34	June Almeida, Bernhard Cinader y Allan F. Howatson	70	Richard M. Goldstein
35	Louis W. Labaw y David R. Davies	71	Allen Beechel
36-42	George V. Kelvin	72	William K. Hartmann Planetary Science Institute
45	Anne d'Harnoncourt	73-77	Allen Beechel
46-47	Robert Sekuler y Eugene Levinson	78	National Aeronautics and Space Administration
48-49	Jerome Kuhl	82-96	A. Caballero, F. Gil y M. Berbel
50-51	Robert Sekuler y Eugene Levinson	98	M. C. Escher
52-53	Jerome Kuhl	99	Andrew Christie
56-57	C. C. Veerman		

Dep. legal: B. 38.999-76
Fotocomposición: Fort, S.A.
Rosellón, 33 - Barcelona-15
Cayfosa. Santa Perpetua de Moguda. Barcelona.
Printed in Spain - Impreso en España

Cartas

Señor Director:

En el número de noviembre de *Investigación* y *Ciencia* dedicado a la alimentación y a la agricultura, el artículo editorial de Sterling Wortman es un tanto desolador. Afirma que debería concentrarse la atención y el dinero sobre procesos y problemas del campo mismo si queremos ganar la batalla del hambre en todo el mundo. Parece ir más lejos al decir que deberíamos concentrar nuestra atención sólo en dichos problemas si se quiere ganar la partida.

No tengo nada que objetar contra la información muy bien presentada por el Dr. Wortman, especialmente contra su adhesión a "los dos componentes de la solución al problema de la alimentación: incremento de la producción de alimentos y aumentos generales de los ingresos familiares, especialmente entre los pobres". Hay, sin embargo, un tercer componente que él y otros muchos agrónomos siempre pasan por alto, y me pregunto por qué razón. El Dr. Wortman describe meticulosamente tres falsas soluciones del problema: exportación continuada de alimentos por parte de las naciones desarrolladas, tecnología agrícola según el patrón occidental y alimentos sintéticos. Ofrece un número de adelantos significativos y esperanzadores para los dos componentes que ha mencionado, enunciando los ocho adelantos más significativos. Estos pueden agruparse así: aumento posible de la producción de las explotaciones, mejor y más generalizada comprensión del problema, disponibilidad mundial de fertilizantes químicos y existencia de instituciones apropiadas a nivel nacional e internacional.

El gran obstáculo surge de la combinación de los siguientes factores: (1) carencia en muchos lugares de medios (sistemas) de entrega del fertilizante, agua y semilla, asesoramiento de expertos y otros presupuestos en cantidad y calidad justas, desde el lugar en donde se hallan disponibles hasta aquel en que se necesitan; (2) carencia en muchas regiones de los adecuados medios (sistemas) para la entrega de los productos de la explotación a la población, y (3) la ineficacia de ambos sistemas, antes y después de la labor agrícola, lo que implica un gran problema dadas las pérdidas que acarrean una y otra deficiencia. Estos tres puntos son puras descripciones simplificadas del gran problema de los sistemas, que se sitúa extramuros de la explotación propiamente dicha.

Estoy de acuerdo en que es necesario y deseable lograr un incremento en la productividad de la explotación, así como un aumento de los ingresos de las familias campesinas. Sin embargo, nada de ello pondrá más alimentos en el plato de más gente en una gran fracción de la tierra, a menos que se preste atención a los problemas que surgen antes del ciclo de trabajo en el campo y una vez han entregado sus productos, o sea, a la cadena de actividades recíprocas y sincronizadas que constituyen propiamente el problema de la alimentación...

H.E. HOELSCHER

Profesor de la
Escuela de Ingeniería,
Universidad de Pittsburgh

Señor Director:

Las observaciones del profesor Hoelscher son justas, y en especial las que se refieren a la debilidad de (1) los sistemas de entrega de suministros y de información y (2) los sistemas mercadológicos ampliamente entendidos, de suerte que comprendan el transporte, almacenamiento, tratamiento y entrega del producto a los consumidores. En definitiva, la mutación del tipo de agricultura de subsistencia a otra altamente productiva exige que todo el sistema -antes y después de la transformación en el campo, como especifica el profesor Hoelscher– funcione adecuadamente. Ningún agricultor puede aumentar su producción, ni lo hará aun pudiéndolo, a menos que se cumplan plenamente cuatro grupos de exigencias como sugerí ante los distintos comités del Congreso de los Estados Unidos (véase mi opúsculo editado por la Rockefeller Foundation: *La situación mundial de la alimentación. Una nueva iniciativa*).

"Los agricultores, incluidos también los que carecen de formación y los que poseen pequeñas propiedades, adoptarán nuevos sistemas siempre y cuando se cumplan cuatro condiciones:

1. Deben contar con sistemas de cultivo más productivos y rentables. Deben ser sistemas completos. En lo referente a cosechas, ello significa variedades apropiadas, mayor abonado, control de enfermedades e insectos, y sucesión adecuada de cosechas. El fomento de tales sistemas implica un trabajo complicado que pueden realizar científicos preparados, pero que sobrepasa la capacidad del "especialista en extensión", cuya formación suele ser insuficiente.
2. Los insumos indispensables –fertilizantes, semillas, pesticidas y créditos– deben ser accesibles al agricultor cuando y donde los necesite, y a un precio razonable. Los sistemas de distribución de estos productos en las zonas rurales deben estar en su lugar y funcionar bien.
3. Al agricultor hay que mostrarle, generalmente en su propio terreno o en el de su vecino, como utilizar eficazmente el nuevo sistema técnico.
4. El agricultor debe saber, antes de invertir en nuevas plantaciones u otras operaciones, que después de la cosecha habrá un mercado para sus productos a un precio fiable, ya que el pequeño propietario no puede aceptar grandes riesgos. Un "mercado" exige carreteras, transporte, demanda efectiva de los productos, precios favorables...

"Se necesitan campañas bien organizadas que fuercen el paso del desarrollo agrícola para que adquiera un ritmo que pocas naciones han experimentado antes. Los elementos clave de tales campañas son la aportación de tecnología biológica y capital para construir la infraestructura que ha de sostener el desarrollo rural. Lo que hay que arbitrar ahora son campañas combinadas en el ámbito rural: la enseñanza de nuevas técnicas y la disponibilidad de nuevas variedades de cosechas y animales, la construcción de una red viaria y de un plan energético, la aportación de fertilizantes, insecticidas y vacunas contra enfermedades de los animales y la disponibilidad de créditos a la exploración y venta de los productos agrícolas. Quienes presten asistencia técnica se verán obligados a crear sistemas de desarrollo agrícola y rural para multitud de gentes inteligentes, pero faltas de educación e incapaces por tanto de emprender sin ayuda externa las innovaciones exigidas en toda explotación agrícola".

Al coincidir con el profesor Hoelscher, no quiero dejar a entender que las mejoras en las explotaciones sean menos importantes de lo que se indicaba en mi

artículo de *Investigación y Ciencia* del mes de noviembre. En los países agrarios los avances de la productividad, unidos a un buen incremento en los ingresos, son requisitos previos al desarrollo económico rural y general.

STERLING WORTMAN

Vicepresidente de la
Fundación Rockefeller,
Nueva York.

Señor Director:

El artículo de Drell y von Hippel ("La guerra nuclear limitada", por Sidney D. Drell y Frank von Hippel; *Investigación y Ciencia*, enero, 1977) fue una reseña adecuada de la situación actual en el tema general de la guerra nuclear limitada estratégica. Lamento que en el título del artículo no figurase la palabra estratégica, por cuanto hay otros planteamientos que han fundado la disuasión de la agresión en el empleo selectivo y limitado de armas nucleares *tácticas* con rendimientos singulares más de 1000 veces menores que los examinados en el artículo de Drell y von Hippel.

Lo que me preocupó en el artículo fue que, después de un amplio examen del tema, no aporte nada sobre lo que podría hacerse. En cierto modo, me pareció que el lector podría suponer que lo que se estaba sugiriendo era un programa de defensa civil a fin de reducir el número de bajas, y no una manera de mantener muy elevado el umbral estratégico.

Puesto que hemos logrado limitar los despliegues masivos de misiles antibalísticos (ABM), sugeriría que quizá sea lo adecuado conseguir un tratado que prohíba todo programa masivo de defensa civil. Si es cierto que la Unión Soviética carece de un programa activo, y es evidente, que tampoco lo tiene Occidente, debería ser fácil proscribir algo que no existe.

H. M. AGNEW

Presidente del Consejo del
Committee on Arms
Control and Disarmament
Washington D.C.

Señor Director:

Lamentamos que al Sr. Agnew le parezca que nuestro artículo "no aporte nada sobre lo que podría hacerse". Nuestro análisis señalaba que un conflicto nuclear limitado, que consistiera en ataques militarmente eficaces contra las fuerzas estratégicas de Estados Unidos, ocasionaría grandísimas bajas civiles, aunque se dispusiera de una amplia protección de defensa civil. Nuestra conclusión explícita fue que la preparación para guerras limitadas de contragolpe nuclear en todos sus aspectos, incluyendo la creación y despliegue de una nueva fuerza de misiles proyectados específicamente como destructores de silos reforzados y la inauguración de nuevos y extensos programas de defensa civil y evacuación de población, no aportaría nada en lo referente a la reducción del "equilibrio de terror" actual y sería perjudicial para el equilibrio estratégico de hoy entre Occidente y el bloque comunista. Proponíamos, por tanto, que ambos frentes debían "limitar su tradicional competición en la tecnología bélica deteniendo la prueba y la puesta en servicio de nuevas armas destinadas a destruir silos reforzados de ICBM (misiles balísticos intercontinentales)". Un paso particularmente directo y verificable para lograr ese fin es negociar, en las conversaciones sobre limitación de armas estratégicas, una cuota anual de disparo de misiles de prueba igual para cada bloque.

En su carta, el Dr. Agnew suscita también otro tema, no examinado en nuestro artículo, el del posible valor disuasorio del *uso* limitado de armas nucleares tácticas de bajo rendimiento para el campo de batalla, especialmente de las "minibombas" que se están proyectando en Los Alamos Scientific Laboratory, que él dirige. Por pequeño que sea su rendimiento, un arma nuclear es inequívocamente nuclear. Una vez iniciada una guerra, la información sobre lo que esté ocurriendo será ambigua y a menudo errónea, y el dominio y el control de los sucesos serán incompletos, tal como lo comprobamos hace poco en Vietnam. Una vez que se empleen en guerra armas nucleares, será muy difícil, si no imposible, verificar los rendimientos, tamaños, números y tipos de las explosiones nucleares desatadas. Cuando se ha cruzado el umbral nuclear nos enfrentamos con el grave peligro de la posible escalada desde una guerra nuclear inicialmente limitada a un holocausto nuclear global.

En resumen, hay sólo un hecho técnico sin ambigüedad, a saber, si se han usado o no armas nucleares. Por consiguiente, para Occidente lo más sensato es adoptar como política nacional el umbral nuclear más elevado posible.

SIDNEY D. DRELL

Stanford University
Stanford, California

FRANK VON HIPPEL

Princeton University
Princeton, N.J.

Los autores

CHRISTOPHER TIETZE y SARAH LEWIT ("El aborto legal"), marido y mujer, han escrito, en común o por separado, más de 200 artículos, la mayoría de ellos sobre el tema de la fertilidad humana y su regulación. Tietze, que es natural de Austria y obtuvo su M.D. por la Universidad de Viena, se trasladó a los Estados Unidos en 1938. Ha trabajado en la Johns Hopkins University School of Hygiene and Public Health, en el Departamento de Estado y en el National Committee on Maternal Health; actualmente en miembro asesor del Population Council's Institute of Policy Research. Lewit, graduada por el Brooklyn College, trabajó en el National Vital Statistics Center y en el National Committee on Maternal Health hasta 1971, en que ingresó en el Population Council como adjunto de investigación.

GLOVER B. TRIPLETT, Jr., y DAVID M. VAN DOREN, Jr. ("Agricultura sin labranza") son agrónomos del Ohio Agricultural Research and Development Center. Triplett, que procedía de la Universidad de Mississippi, se doctoró en agronomía en Michigan en 1959, un año más tarde que Van Doren, quien, a su vez, procedía de la Universidad de Illinois, y se había especializado en edafología. Tras acabar sus estudios se integran en el centro experimental citado, donde siguen investigando en nuestros días. Para el desarrollo de sistemas que prescinden de la labranza emplean sus especialidades complementarias: Triplett en el control de las malas hierbas y Van Doren en la física de los suelos.

S. W. HAWKING ("La mecánica cuántica de los agujeros negros") enseña física teórica en la Universidad de Cambridge. Nació en Oxford en 1942 y se graduó en su famosa universidad en 1962. Realizó su tesis de doctorando sobre relatividad general en Cambridge, bajo la dirección de D. W. Sciama. En 1974-1975 fue Sherman Fairchild Distinguished Scholar en el Instituto de Tecnología de California. Entre otras numerosas distinciones ha recibido la Eddington Medal, de la Royal Astronomical Society, y el Premio Heineman de Física matemática, concedido por la American Physical Society y el American I. of Physics.

J. DONALD CAPRA y ALLEN B. EDMUNDSON ("El lado de unión de los anticuerpos") son inmunólogos en dos instituciones médicas en el suroeste de los Estados Unidos. Capra enseña microbiología en el Health Science Center, de la Universidad de Texas, en Dallas. Antes, había trabajado como investigador en el National Institute of Health, en la Rockefeller University y en la Mount Sinai School of Medicine de Nueva York. Edmundson es profesor de bioquímica en el Medical Center de la Universidad de Utah.

ROBERT SEKULER y EUGENE LEVINSON ("La percepción de objetos en movimiento") pertenecen al claustro de la Northwestern University, en donde Sekuler es profesor y coordinador del departamento de psicología y, Levinson, colaborador de postdoctorado y profesor invitado en el centro de ingeniería biomédica de la universidad. Sekuler se graduó en la Universidad de Brandeis y se doctoró en la de Brown. En 1965 entró a formar parte de la facultad de la Northwestern, después de haber pasado un curso de postdoctorado en el Instituto de Tecnología de Massachusetts. Su colaboración con Levinson en varios estudios sobre la percepción visual comenzó en 1971, cuando éste asistía a un cursillo para graduados que daba Sekuler sobre la percepción.

ERNEST RABINOWICZ ("Exoelectrones") es profesor de ingeniería mecánica en el Instituto de Tecnología de Massachusetts. Se graduó y doctoró por la Universidad de Cambridge. Está en el M.I.T. desde 1950. Como experto en los campos de fricción y fatiga, escribe: "Difícilmente se produce un acontecimiento técnico en el mundo, sea una nueva prueba espacial, un nuevo reactor nuclear, un nuevo sistema computador o un nuevo motor, sin que se cree un problema de fricción o fatiga nuevo".

WILLIAM K. HARTMANN ("Formación de cráteres en el sistema solar") trabaja en Tucson (Arizona) como científico jefe en el Instituto de Ciencia Planetaria, una sección de la compañía Science Applications. Se graduó en física por la Universidad de Pennsylvania; se licenció en geología y doctoró en astronomía en la Universidad de Arizona. Ha investigado gran variedad de problemas relacionados con el origen de los sistemas planetarios y la evolución de la superficie de los planetas. Es autor de *Moons and Planets*, un texto de introducción a la ciencia planetaria, y tiene en preparación un nuevo texto sobre astronomía general.

ARTURO CABALLERO, FRANCISCO GIL y MIGUEL BERBEL ("La productividad de prados mediterráneos") trabajan en la Universidad de Barcelona. Arturo Caballero es catedrático de fisiología vegetal de la Facultad de Biología y director del Departamento y de los Campos Experimentales de la misma Facultad. Miembro de la Real Academia de Ciencias y Artes de Barcelona. Su línea de investigación actual se centra en la ecofisiología, sobre todo, en relación con especies pratenses. Francisco Gil, licenciado y doctor en Biología (1974) por Barcelona, se integra en dicha Universidad en 1971. Su campo de actuación se centra en el estudio de modelos matemáticos y computacionales sobre aspectos paramétricos y temporales de los organismos vegetales y en el contexto de la nutrición mineral. Miguel Berbel, perito agrícola por la Escuela de Peritos de Barcelona (1963), colabora desde entonces en el mismo Departamento, como técnico de Campos Experimentales, habiéndose especializado en el tratamiento estadístico de datos mediante computadoras electrónicas.

ERRATA

Número 4, enero de 1977

Pág.	*columna*	*línea*	*dice*	*debe decir*
28	1	43	*(véase la ilustración de la pág. 33)*	*(véase la ilustración de la pág. 29)*
29	1	24	Esto significa	Esto no significa
30	1	54	*(véase la ilustración de la pág. 33)*	*(véase la ilustración de la pág. 32)*

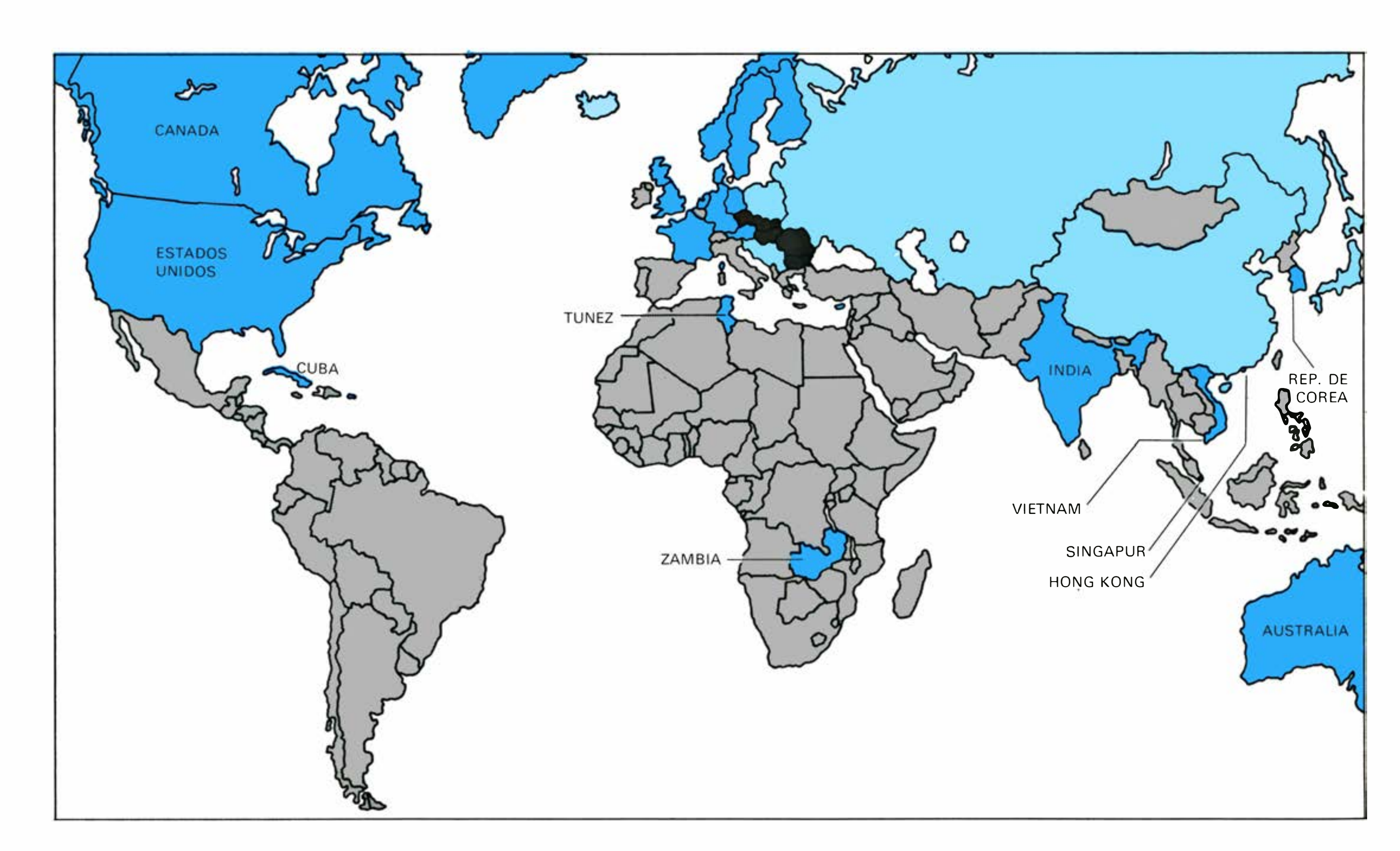
El aborto legal

La tendencia hacia la legalización del aborto provocado ha dado como resultado nuevos estudios sobre su práctica, una mayor información al respecto, mejores técnicas médicas y menores índices de mortalidad

Christopher Tietze y Sarah Lewit

Hace ocho años escribíamos en *Scientific American* que algunas naciones habían legalizado el aborto provocado y que, como consecuencia de ello, su práctica "está emergiendo de las sombras y se está convirtiendo en tema de discusión y controversia en todo el mundo". Hoy está claro que aquello que observábamos entonces no era más que el comienzo de una tendencia que ha ido generalizándose. A lo largo de los últimos diez años, en muchos países, el aborto ha dejado de ser una práctica en la mayoría de los casos vergonzosa para convertirse en una práctica médica aceptada; de ser un tema de murmuración ha pasado a ser objeto de un debate público abierto. Se está lejos de llegar a un acuerdo sobre la cuestión. El debate se caracteriza por unas diferencias de opinión honradas y profundamente sentidas, y las políticas nacionales con respecto al aborto todavía están sujetas a cambios en cualquier sentido. Por lo general, no obstante, los progresos de la década han intensificado la conciencia pública sobre un asunto de enorme significación humana, han permitido mejoras en las técnicas médicas y han reducido el número de muertes innecesarias.

La acción legislativa ha transformado drásticamente la situación del aborto. Desde 1966 muchos países han modificado unas leyes restrictivas para llegar a legalizar el aborto a petición de la mujer embarazada, o bien para especificar una



LAS LEYES Y LAS PRACTICAS ABORTIVAS se han vuelto menos restrictivas durante la pasada década en más de veinte países *(azul oscuro).* **En otros, con leyes liberalizadoras o moderadamente restrictivas, no se han producido cambios significativos** *(azul claro).* **Cua-**

amplia lista de indicaciones médicas y socioeconómicas para la interrupción legal de un embarazo no deseado. Estos cambios legales, por otra parte, han contribuido a modificar las actitudes de los profesionales. Como respuesta, en parte, a estos cambios de actitud, los médicos han desarrollado nuevos y más seguros métodos que posibilitan la práctica de un aborto en los inicios del embarazo, sin necesidad de internar a la paciente ni de recurrir a la cirugía mayor, y que reducen el período de hospitalización en los casos de embarazo más adelantado. La creciente experiencia en la práctica del aborto legal, y el mayor conocimiento de las consecuencias del aborto ilegal en la salud pública, han animado a los organismos sanitarios a recoger y publicar estadísticas oficiales de las que antes no podía disponerse. La necesidad de establecer unas líneas de orientación para los médicos y hospitales que practican un gran número de abortos ha impulsado la realización de estudios sobre las posibles complicaciones del aborto. Los resultados de estos estudios, ampliamente difundidos, han animado a la mujer que desea abortar a hacerlo lo antes posible y en unas condiciones médicas de total seguridad.

No es menos cierto que el interés oficial y público por la libertad del aborto que se da en muchos países se ve superado por la indiferencia en amplias áreas del mundo. El resultado es que la información sobre el aborto a nivel mundial sigue siendo sumamente insatisfactoria. En muchos países, especialmente en aquellos que tienen una legislación restrictiva, los abortos ilegales siguen realizándose en elevado número, cuyas cifras no se recogen y, generalmente, se desconocen. Sumando los datos nacionales, de precisión variada, se puede estimar que cada año se provocan en el mundo de 30 a 55 millones de abortos. La cifra menor significa una relación aborto-nacimiento del orden de uno por cada cuatro, y la cifra mayor significa una relación de uno por cada dos. Quizá la mitad de estos abortos son realizados legalmente; es imposible estar seguros porque incluso algunos países con leyes realmente permisivas, entre ellos la URSS y China, o no disponen de cifras exactas o prefieren no hacerlas públicas.

Un examen de la legislación a lo largo de la pasada década muestra que más de una veintena de países han llevado a cabo cambios significativos en el sentido de disminuir la restricción y, en muchos casos, han escogido la total legalización *(véase la ilustración de las dos páginas siguientes)*. Cuatro países de la Europa oriental, que tenían unas leyes sobre el aborto muy permisivas, las han modificado en sentido contrario. Otros países que tenían leyes no restrictivas o sólo moderadamente restrictivas no han realizado cambio alguno en el curso de la década. En los países restantes, la mayoría de ellos de Latinoamérica y Africa, no se han producido cambios significativos en las leyes que o bien prohíben el aborto sin excepción o sólo lo permiten para prevenir una seria amenaza para la vida o la salud de la madre.

Alrededor de un tercio de la población mundial vive en estos momentos en países con leyes sobre el aborto no restrictivas, o sea, países en los que los embarazos pueden interrumpirse a voluntad como mínimo hasta un determinado tiempo de gestación (de 10 semanas hasta 24, siendo 12 semanas el límite máximo más común) y, después de ese tiempo, por indicación médica; en algunos países, la concesión del permiso para abortar a voluntad propia puede depender de la edad de la mujer y del número de hijos.

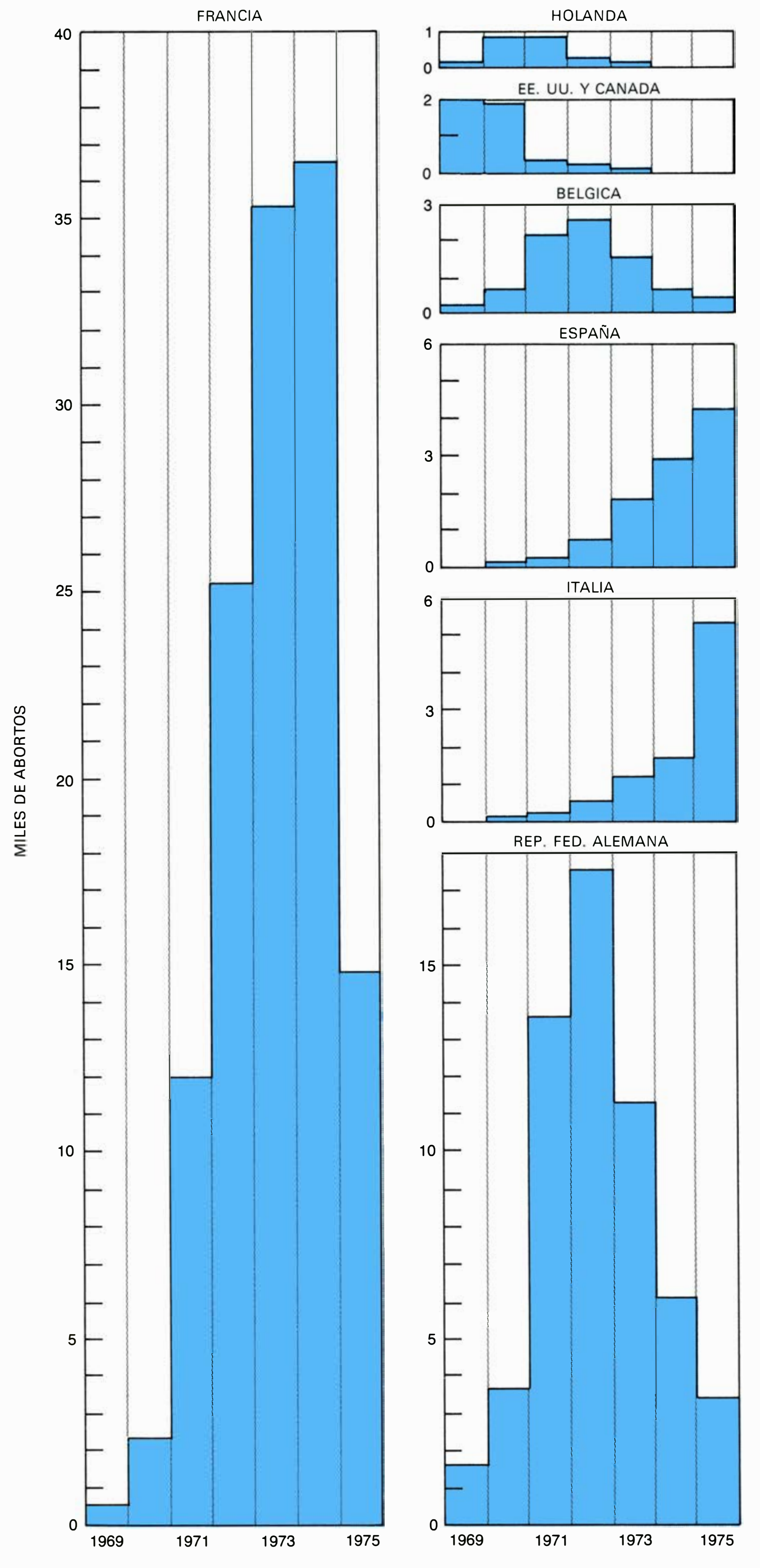
Otro tercio de la población mundial vive en países con leyes sobre el aborto moderadamente restrictivas, en los que los embarazos no deseados pueden interrumpirse no a simple petición de la interesada sino después de una serie de razones médicas, psicológicas y socioeconómicas ampliamente estudiadas. La vivienda, los ingresos, la categoría del marido y otros factores que afectan a la situación de la mujer pueden entrar en consideración. Las razones jurídicas incluyen generalmente actos criminales como violación o incesto. Los avances médicos en las técnicas para la diagnosis precoz de las anormalidades fetales han suministrado nuevas razones para permitir el aborto en muchos países. Una disposición de la British Abortion Act, de 1967 (y algunas otras leyes que tienen a ésta como modelo), permiten el aborto si la continuación del embarazo puede suponer un riesgo mayor que su interrupción.

El tercio restante de la población mundial vive en países en los que el aborto o bien es absolutamente ilegal o bien está permitido sólo en el caso de que la vida de la mujer o su salud se vean seriamente amenazadas por la continuación del embarazo.

La realidad no siempre coincide con



tro países han pasado de unas regulaciones menos restrictivas a otras más restrictivas *(gris oscuro)*. **En la mayoría, el aborto sigue siendo restrictiva, cuando no desconocida** *(gris claro)*.



las leyes, desde luego. En bastantes países en los que está prohibido el aborto pueden, sin embargo, recurrir a él quienes tengan dinero para pagarlo. Por otro lado, en países con leyes no restrictivas médicos conservadores, directores de hospitales o funcionarios pueden negarse a realizarlos o no permitir que se realicen. La falta de facilidades o de servicios apropiados y a precios razonables puede impedir también que una mujer aborte, según desea.

En muchos casos, mujeres que no pueden interrumpir la gestación en su propia comunidad viajan a lugares con leyes o normas menos restrictivas. El récord en el número de abortos alcanzado en Inglaterra después de la promulgación de la Abortion Act, en 1967, es ilustrativo al respecto *(véase la ilustración de la página 9).* Un número considerable de mujeres de América y del continente europeo fueron a Londres a abortar a partir de 1968, pero las cifras por países descendieron bruscamente en los casos en que el aborto se volvió asequible en el propio país o en los países vecinos. Por ejemplo, la aprobación de una ley no restrictiva en el estado de Nueva York, en 1970, se vio reflejada por un descenso al año siguiente en el número de mujeres norteamericanas y canadienses que abortaron en Inglaterra. La aparición de clínicas holandesas en las que se practica el aborto con tolerancia oficial redujo el número de pacientes de ese país en 1972 y el número de pacientes de los países vecinos, Bélgica y Alemania occidental, un año después. La aprobación de una ley no restrictiva en Francia redujo, en un 60 por ciento en un año, el número de mujeres francesas que iban a Londres a abortar. Por otro lado, un número creciente de mujeres seguían acudiendo en 1975 procedentes de Irlanda, Italia y España, países en los que el aborto sigue estando prohibido.

En los Estados Unidos, los cambios en los distintos estados relativos a las

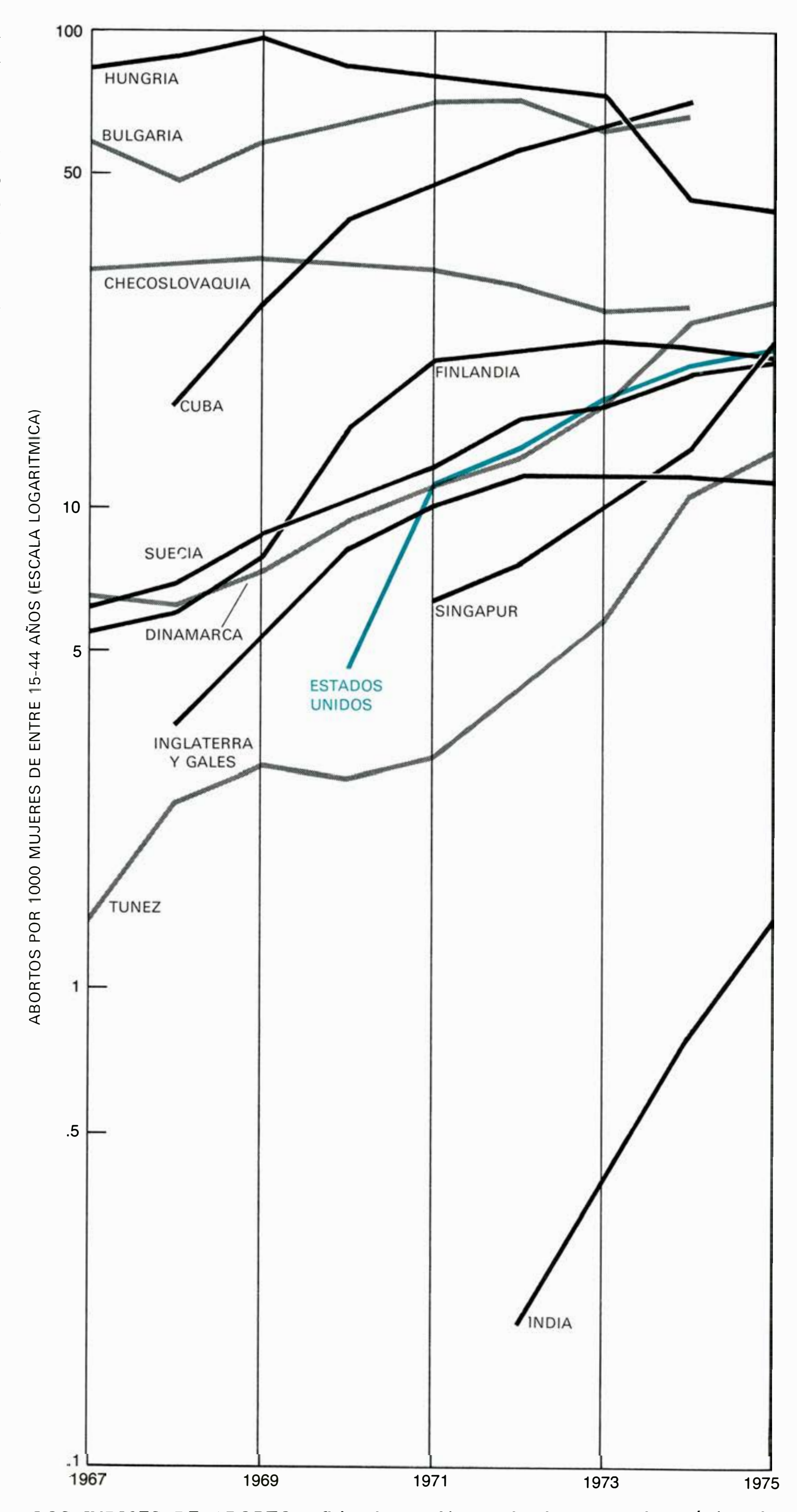
EL NUMERO DE NO RESIDENTES que consiguieron abortar en Inglaterra después de que fue permitido en 1968 tendió a incrementarse, al ir conociendo las mujeres la posibilidad de esta opción, pero descendió sensiblemente en los casos en que el aborto se volvía más fácil de conseguir en sus propios países. Los datos reflejan la reducción en las restricciones en el estado de Nueva York (y, con ello, las facilidades para abortar para las mujeres de Estados Unidos y Canadá), en Holanda (cambio que afectó también a las mujeres de Bélgica y de la República Federal Alemana) y, más recientemente, en Francia.

leyes sobre el aborto siguieron en la pasada década una tendencia paralela a la internacional. Se inició en 1967 con la promulgación, en una docena de estados, de legislaciones basadas en el modelo de código penal del American Law Institute. Esas leyes permitían el aborto en aquellos casos en que el embarazo supusiera un "riesgo sustancial" para la salud mental o física de la madre, o el niño pudiera nacer con "un grave defecto físico o mental", o si el embarazo era resultado de una violación o de un incesto. En 1970, cuatro estados fueron más allá: aprobaron unas leyes que permitían el aborto sin necesidad de especificar razón alguna, con lo que, en realidad, autorizaban el aborto voluntario.

Dos fallos del Tribunal Supremo, en 1973, invalidaron las leyes restrictivas sobre el aborto que todavía estaban vigentes en la mayoría de los estados al sentenciar, basándose en gran parte en razones de respeto a la intimidad, que no podría negarse el aborto a una mujer en los tres primeros meses (primer trimestre) del embarazo. Después de los tres meses, dictó el Tribunal, un estado "puede regular el procedimiento relativo al aborto de modo que vele razonablemente por la salud materna". Una vez que el feto ha alcanzado viabilidad (la capacidad de supervivencia en condiciones favorables), lo que normalmente sucede hacia las 24 semanas, el estado puede "prohibir el aborto, salvo si es necesario [...] para la conservación de la vida o la salud" de la mujer.

Estas sentencias del Tribunal Supremo nunca han sido llevadas a la práctica enteramente. Un vigoroso movimiento de defensa del "derecho a la vida" ha venido ejerciendo con éxito una fuerte presión en los cuerpos legislativos de algunos estados y han conseguido que se promulgaran leyes restrictivas y que muchos directores de hospitales prohibieran la realización de abortos en contra de aquellas sentencias. Algunos estados empezaron a exigir la autorización paterna en los casos de menores de edad o, la marital, en el caso de la mujer casada.

Como mínimo en un estado, el método preferente para realizar el aborto en el segundo trimestre fue declarado ilegal. En los hospitales católicos no se realizan abortos y un gran número de hospitales no católicos se muestran poco dispuestos a montar, siquiera fuera, las instalaciones necesarias para su realización. Algunas de esas leyes estatales restrictivas ya han sido recusadas ante los tribunales y han sido declaradas anticonstitucionales. Los intentos por mo-



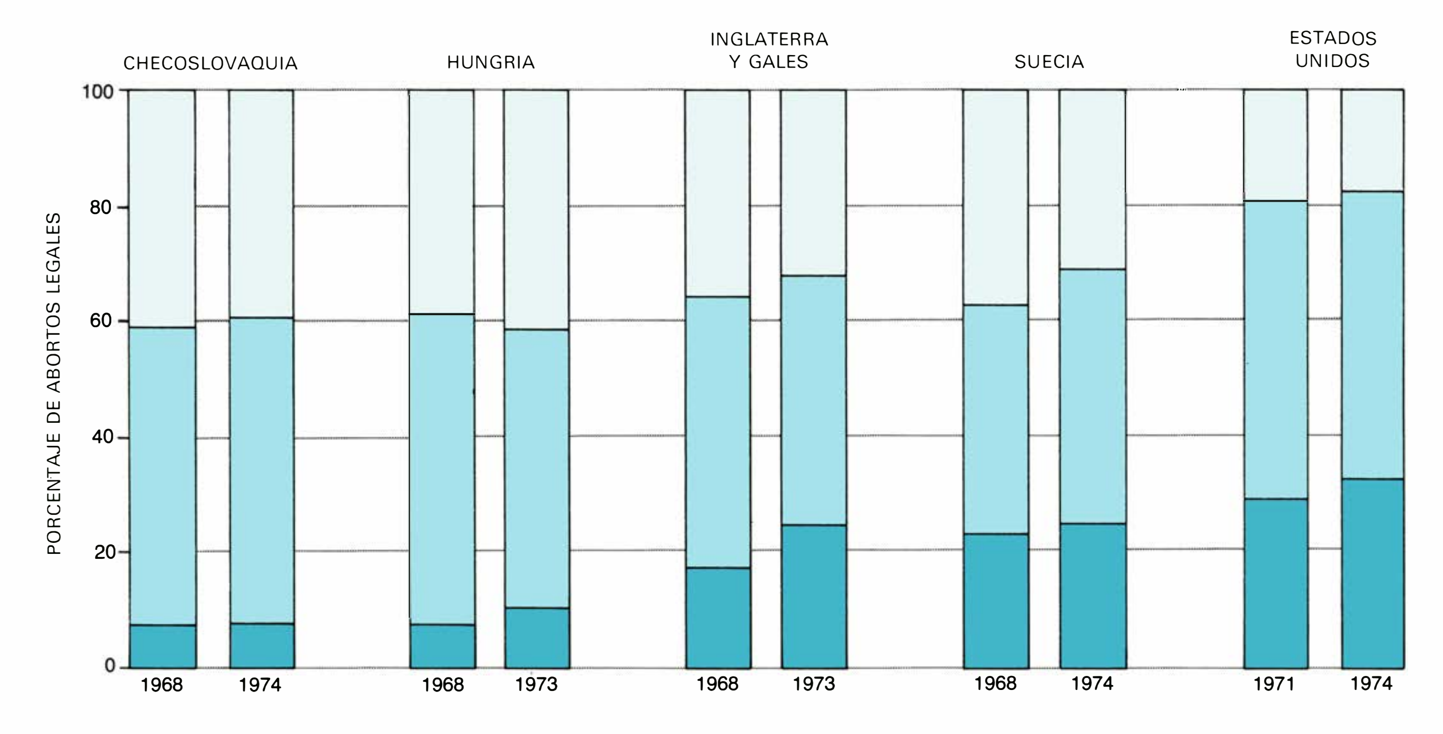
LOS INDICES DE ABORTO reflejan los cambios en las leyes y en las prácticas, la disponibilidad de servicios, la extensión y la eficacia de la anticoncepción y el grado de correlación entre aborto y regulación de la natalidad. En el presente gráfico, el número de abortos legales por cada 1000 mujeres, de edades comprendidas entre los 15 y los 44 años, se ha fijado sobre una escala logarítmica; es decir, las líneas paralelas expresan índices de cambio iguales. Los datos relativos a Inglaterra y Gales se refieren solamente a residentes.

dificar la Constitución federal, bien en el sentido de que declare ilegal el aborto, bien para que permita que cada estado pueda declararlo, han recibido cierto apoyo pero no han pasado todavía del comité del Congreso. El anterior Congreso, no obstante, agregó una cláusula a un proyecto de ley presupuestaria en la que se prohibía al Gobierno Federal la ayuda financiera a los abortos realizados a través de Medicaid, lo que provocó su denuncia inmediata como una negación anticonstitucional de la igualdad de derechos para los pobres.

La relajación en las barreras legales que impiden el aborto y los cambios en las actitudes de la opinión pública han contribuido a vencer la resistencia que oponían algunos países a facilitar información sobre el aborto. Hoy existen estadísticas por naciones sobre unos tres millones de abortos legales al año, con muchas diferencias de un país a otro en lo relativo a su amplitud y a la cantidad de detalles suministrados.

Nosotros hemos calculado el índice de abortos legales (por cada 1000 mujeres de 15 a 44 años) en 12 países con leyes no restrictivas o moderadamente restrictivas y para los que se disponía de datos que abarcaban todo, o casi todo, el decenio comprendido entre 1966 y 1975. En la mayoría de estos países, el índice de abortos aumentó a lo largo de la década. Una razón para este crecimiento esté, probablemente, en un cambio –cuyo alcance todavía no puede acreditarse– en el sentido de haber pasado de unos procedimientos ilegales y clandestinos a otros legales y registrados. Además, algunas mujeres que no habrían aceptado un aborto ilegal se sienten libres para escoger un procedimiento legal cuando éste está a su disposición. Las curvas de los índices de abortos muestran fluctuaciones en las políticas nacionales sobre el asunto, pero unas políticas más o menos parecidas pueden tener efectos distintos (*véase la ilustración de la página 9*). En Singapur, por ejemplo, el cambio de una ley moderadamente restrictiva a una ley no restrictiva, en 1974, tuvo como consecuencia un 65 por ciento de aumento en el índice de abortos. En Hungría, la aprobación de unas disposiciones restrictivas en 1973 aceleró bruscamente el descenso que venía observándose desde el punto máximo alcanzado en 1969; en Bulgaria y Checoslovaquia unas disposiciones en el mismo sentido produjeron un descenso inicial que fue seguido de un leve ascenso. Cuando se alcanza un determinado nivel en un país con leyes no restrictivas, como Inglaterra, Gales o Finlandia, éste refleja probablemente el incremento y una práctica más eficaz de la anticoncepción. Un índice muy bajo, como el de la India (1,4 en 1975), puede atribuirse en parte a la falta de registro de los mismos abortos legales, pero también es cierto que la falta de médicos y de facilidades ha limitado el número de abortos legales practicados en ese país. En algunos países con índices especialmente elevados, las mujeres, al parecer, confían más en el aborto para impedir nuevos nacimientos, mientras que en otros lo que ocurre es que se practica amplia y eficazmente la anticoncepción, limitando los embarazos y, por consiguiente, el número de abortos.

En cinco países en los que pudo compararse la distribución por edades de las mujeres que han recurrido al aborto legal al principio y al final de la década, se advirtió que el número de mujeres que interrumpieron la gestación fue mayor entre las que contaban entre 20 y 30 años que entre las menores o mayores. Actualmente, el grupo de las jóvenes (de 13 a 19 años) constituye la tercera parte de la totalidad de los abortos practicados en Estados Unidos y alrededor de la cuarta parte en Inglaterra y Gales y en Suecia. En Checoslovaquia y Hungría la proporción de mujeres muy jóvenes es mucho menor: el análisis de los índices específicos por edades muestra claramente que se dan índices más elevados entre las mujeres adultas de estos dos países que entre las jóvenes de los restantes. En las cinco naciones se realizaron más abortos entre mujeres de menos de 20 años hacia



LA PROPORCION DE MUJERES MENORES de 20 años que abortaron en Inglaterra y Gales, Suecia y Estados Unidos es mayor que la proporción correspondiente a Hungría y Checoslovaquia. Las barras señalan el porcentaje de abortos legales que se realizaron en los años indicados a mujeres menores de 20 años *(color oscuro)*, de entre 20 y 29 años *(color intermedio)* y a mujeres de 30 o más años *(color claro)*.

el final de la década que hacia el principio. Esto puede ser el reflejo de una maduración física más temprana, un matrimonio más tardío (que aumenta la probabilidad de embarazos prematrimoniales), una distancia creciente entre la madurez sexual y la social, una mayor aceptación del aborto como una alternativa al matrimonio forzado o a los nacimientos fuera del matrimonio, una liberación de las leyes sobre el aborto y los cambios en la actitud y en la práctica por parte de la clase médica.

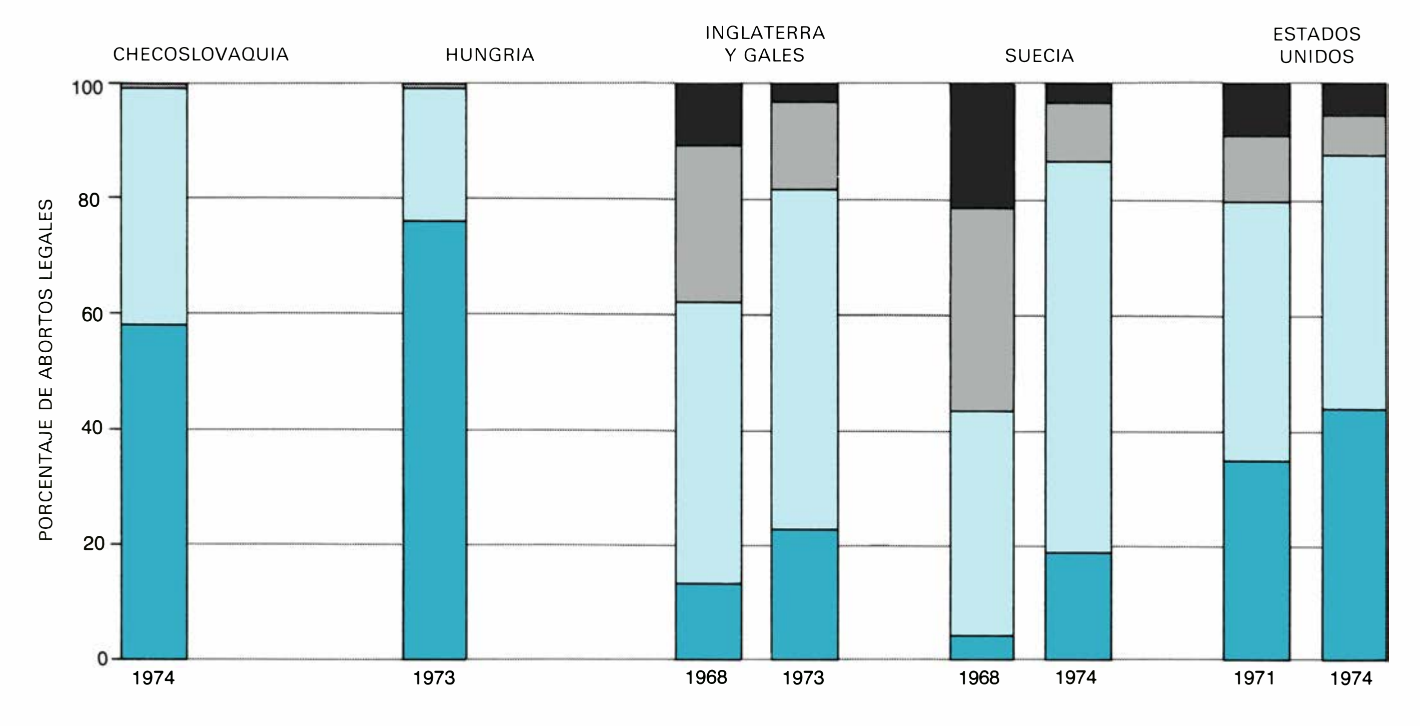
Existe poca información digna de crédito respecto al número de mujeres que se han hecho practicar más de un aborto a lo largo de su vida. Según parece, cerca de un 60 por ciento de las mujeres húngaras que abortaron voluntariamente en 1973, lo habían hecho como mínimo una vez anteriormente. Esta cifra fue, para la ciudad de Nueva York, en 1974, del 22 por ciento. Estos tantos por ciento crecerán inevitablemente con el tiempo, incluso allí donde se practica concienzudamente la anticoncepción, a menos que se incremente el ofrecimiento y la aceptación de la esterilización por parte de las mujeres (y los hombres) que no desean más hijos.

La contribución más importante de los últimos avances en el campo del aborto es la creciente comprensión de que los abortos precoces son más seguros y menos traumáticos. El aborto en el primer trimestre de gestación suele realizarse en la actualidad por aspiración en el útero más que por raspado (legrado). Cuando el embarazo está más adelantado, la técnica médica tradicional ha sido ampliamente reemplazada por la inyección en el útero de una sal o solución hormonal que provoca la expulsión del feto. En tres países para los que se dispone de cifras fidedignas sobre la distribución de los abortos por períodos de gestación, la proporción de abortos en los tres primeros meses ha aumentado al comprender médicos y mujeres su relativa inocuidad y al dar mayores facilidades los servicios legales respectivos. (En Checoslovaquia y Hungría siempre se han practicado muy pocos abortos legales en el segundo trimestre de gestación, y sólo están autorizados por razones médicas.)

Pese a la bien probada mayor seguridad de los abortos precoces algunas mujeres siguen sometiéndose a ellos cuando su gestación está más avanzada y el procedimiento tiene mayores riesgos. Un estudio realizado en Inglaterra y Gales mostró que el nivel profesional incidía en este hecho: las mujeres que eran trabajadoras no cualificadas (o lo eran sus maridos) tendían a abortar en un tiempo de embarazo más adelantado que las mujeres (o sus maridos) con trabajos cualificados, administrativos o con profesiones liberales. En los Estados Unidos, las mujeres negras suelen abortar más tarde que las blancas.

Las mujeres muy jóvenes tienden también a abortar más tarde. En la ciudad de Nueva York, la cifra más elevada de abortos realizados en el segundo trimestre de embarazo se registró en adolescentes menores de 15 años y la cifra más baja se registró en mujeres de 35 años o más *(véase la ilustración de la página siguiente)*. El hecho de que las mujeres muy jóvenes no recurran antes al aborto puede ser atribuible a una incapacidad para reconocer los síntomas del embarazo, a la creencia de que "esto no puede ocurrirme precisamente a mí" o la idea ilusoria de que "esto pasará", a ignorancia sobre a dónde ir en busca de ayuda, resistencia a confesarlo a los padres o falta de dinero; los engorrosos trámites administrativos también pueden retrasar la obtención de la necesaria autorización en tanto que la gestación sigue su curso.

Algunas mujeres se ven obligadas a abortar en un estado avanzado de gestación: son las que se someten a unas pruebas de diagnóstico por las que llegan a saber que el feto presenta un defecto físico o mental importante. Actualmente se pueden detectar durante el embarazo unas 60 anormalidades fetales. El método de diagnóstico principal es la amniocentesis, la extracción para análisis de



UNA TENDENCIA HACIA EL ABORTO PRECOZ (en especial durante los tres primeros meses) puede observarse en Inglaterra y Gales, Suecia y los Estados Unidos. Las barras expresan el tanto por ciento de abortos en las ocho primeras semanas de gestación *(color oscuro)*, de la 9 a la 12 *(color claro)*, de la 13 a la 16 *(gris claro)* y desde la 17 *(gris oscuro)*. En Checoslovaquia y Hungría son mínimos desde la 12.

una pequeña cantidad del líquido amniótico en que está inmerso el feto. También puede examinarse el feto por medio de ultrasonidos, o todavía en procedimiento experimental, por fetoscopia (inspección visual directa) o también por análisis de la sangre fetal. El momento óptimo para este tipo de pruebas prenatales es alrededor de las 16 semanas de gestación; puede requerirse un tiempo supletorio para el cultivo de tejidos, siempre dentro del período de viabilidad. El número de abortos realizados como consecuencia de este tipo de pruebas sigue siendo muy reducido (menos de 200 al año en Estados Unidos) porque estos procedimientos de diagnóstico sólo pueden llevarse a cabo en unos cuantos centros hospitalarios.

El aborto provocado, en cualquier fase del embarazo, comporta un cierto riesgo de complicaciones, que van desde algunas dolencias de escasa importancia hasta una eventual fatalidad; dependen de las circunstancias y, desde luego, de la fase en que se halle la gestación. Recientes estudios sobre abortos practicados en hospitales y en clínicas muestran que las complicaciones presentadas pueden agruparse en dos categorías principales: precoces: (durante la intervención o durante el mes siguiente) y tardías más de un mes después de la operación). Las primeras son generalmente ligeras molestias, como dolores de cabeza o pequeñas hemorragias. El índice de complicaciones precoces graves es sólo del 5 por ciento para abortos realizados durante el primer trimestre y de un 2 por ciento para los realizados después de este período, según un amplio estudio llevado a cabo en los Estados Unidos durante el bienio 1970-1971. Todas estas complicaciones incluyen perforaciones del útero, hemorragias importantes, laceración del cuello de la matriz (cérvix), graves perturbaciones de coagulación e hipernatremia (exceso de sodio en sangre) debida al aborto provocado por medio de una solución salina, graves infecciones pélvicas o generalizadas, tromboflebitis venosas y embolias pulmonares.

La identificación de las complicaciones tardías es, todavía, muy incompleta. Las mujeres Rh-negativas pueden ser sensibilizadas por las células sanguíneas de un feto abortado Rh-positivo, pero esto puede evitarse por medio de la inmunización con Rh-globulina. El aborto por incisión en el útero (histerotomía) puede conducir a una endometriosis, o crecimiento anormal del tejido mucoso uterino. El riesgo, de que se produzcan, en posteriores embarazos, un aborto espontáneo en el segundo trimestre, un parto prematuro o una muerte precoz del niño ha sido estudiado pero los resultados son todavía conflictivos o no concluyentes. El valor de anteriores estudios referentes a estos efectos es cuestionable ya que las mujeres que tienen un parto normal después de un aborto suelen dar cuenta con menos frecuencia de haberse sometido a un aborto que las que tienen embarazos que acaban de modo no satisfactorio.

La Organización Mundial de la Salud ha patrocinado estudios prospectivos en siete países europeos sobre los efectos, en un posterior embarazo y parto, de variables tales como la duración del embarazo precedente interrumpido, la técnica (raspado o aspiración) empleada y el tiempo transcurrido entre el aborto y la posterior fecundación. Los resultados estarán disponibles a fines de este año.

El papel desempeñado por el aborto provocado en ulteriores casos de esterilidad secundaria, embarazo ectópico (desarrollo del feto fuera del útero), placenta previa (desarrollo de la placenta próximo a la abertura del útero) y menstruación dolorosa no ha sido probado; deben realizarse más investigaciones sobre estos aspectos y sobre los efectos psicológicos (si los hay) del aborto provocado y también del caso contrario, es decir, de dar a luz a un hijo no deseado.

El índice de mortalidad para todos los tipos de aborto (legal, ilegal y espontáneo) ha descendido bruscamente a lo largo de la pasada década, con todo y que el número de abortos legales se ha incrementado con las nuevas facilidades. En Estados Unidos, por ejemplo, se dio un promedio de 5,7 muertes de abortantes al año por cada millón de mujeres cuya edad oscilaba entre 15 y 44 años en el período comprendido entre 1963 y 1967; este promedio fue de una muerte por millón para 1974. Un descenso comparable, partiendo de un nivel comparable también, se ha registrado en Inglaterra y Gales en el mismo período. En Rumanía, por el contrario, el número de muertes debidas al aborto se ha incrementado casi en seis veces desde que una ley liberalizadora sobre el aborto fue substituida por otra restrictiva en 1966.

La reducción de la mortalidad por aborto coincide en muchos países con la tendencia a practicarlo más tempranamente (cuando la intervención es más sencilla, más segura y menos traumática) y con la disminución de la edad media de las mujeres que se someten a él (puesto que, independientemente de la fase de gestación, el riesgo es mayor en mujeres de más edad). Otro factor que probablemente contribuye a este descenso de la mortalidad es el mayor intercambio entre los médicos de los conocimientos que van adquiriendo al ampliar su experiencia a base de nuevas interven-

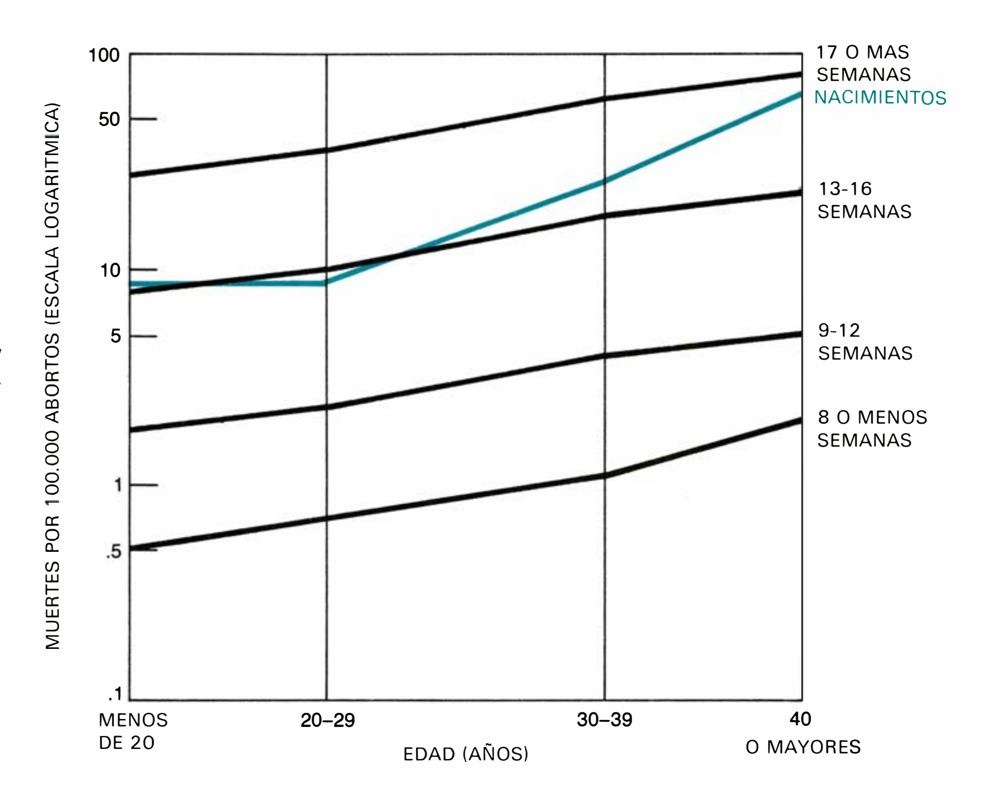
LOS ABORTOS TARDIOS presentan la mayor frecuencia entre las mujeres más jóvenes y, la menor, entre las mujeres de 35 o más años. Las barras muestran qué proporción de intervenciones se llevaron a cabo durante las 12 primeras semanas de gestación *(color)* y durante la treceava semana o más adelante *(gris)*, en la ciudad de Nueva York, en el decurso de 1974.

ciones. Puede establecerse una comparación especialmente interesante entre el riesgo de mortalidad relacionada con el aborto y el mismo riesgo relacionado con el embarazo y el parto. El índice de mortalidad para los abortos legales realizados durante las primeras 12 semanas de embarazo es mucho más bajo que el debido a complicaciones del parto, según cifras relativas a los Estados Unidos y Gran Bretaña *(véase la ilustración de esta misma página)*. Los índices de mortalidad por aborto entre las semanas 13 y y 16 y los de mortalidad debida al parto son similares, excepto para las mujeres mayores, en cuyo caso el parto supone un riesgo mayor. Solamente los abortos practicados a las 17 semanas o más de embarazo presentan un índice de mortalidad generalmente superior al del parto.

En países en los que la anticoncepción se practica amplia y eficazmente, como es el caso de Estados Unidos y Gran Bretaña, los índices de abortos son sustancialmente inferiores a los de los países en que no se practica la anticoncepción, se hace en muy poca medida, o en los que los embarazos no previstos acaban frecuentemente en abortos. Los modernos métodos anticonceptivos, como la "píldora" o el dispositivo intrauterino (DIU), pese a ser altamente eficaces en la evitación de embarazos, presentan algunos efectos secundarios conocidos; estos efectos pueden, en algún caso, ser fatales. Es conveniente, por lo tanto, investigar los riesgos relativos del aborto precoz y de los distintos métodos anticonceptivos como medios para el control de la natalidad.

Uno de nosotros (Tietze) colaboró recientemente en un estudio, basado en un modelo cibernético, sobre la mortalidad que es atribuible en los países desarrollados a los cuatro métodos principales de anticoncepción (píldora, dispositivo intrauterino, preservativo y diafragma) y al aborto durante el primer trimestre. La mortalidad total era la suma de la mortalidad atribuible al sistema de control de natalidad seguido y la atribuible a todo embarazo accidental y a posibles partos.

En el caso de la píldora, la mayor parte de la mortalidad obedecía al método anticonceptivo; para el dispositivo intrauterino, se debió más o menos por igual al propio dispositivo y a embarazos accidentales; para el caso de los inocuos, pero menos eficaces, preservativos y diafragmas, las muertes se debieron en su totalidad a embarazos fortuitos, y, en el caso del aborto precoz, la mortalidad sólo se atribuyó al procedimiento abortivo, ya que el embarazo se había interrumpido.



LA MORTALIDAD ATRIBUIBLE AL ABORTO LEGAL, a distintas edades y en distintas fases de gestación, se compara en el presente gráfico con la mortalidad atribuible al parto. Los datos se refieren a la experiencia conjunta de los EE.UU. e Inglaterra; la escala vertical es logarítmica. El índice de mortalidad por aborto aumenta con el tiempo de embarazo y, en menor grado, con la edad de la mujer. Para todos los grupos por edades, el índice de mortalidad para el aborto practicado en el primer trimestre es inferior al índice correspondiente al parto.

Los resultados mostraron que, entre las mujeres de menos de 30 años, unos índices de mortalidad muy bajos (entre una y dos muertes por 100.000 mujeres y año) iban asociados con cada uno de los métodos anticonceptivos y con el aborto precoz cuando cada uno de estos sistemas de control de la natalidad era utilizado por separado. En comparación, el riesgo debido al embarazo y al parto para mujeres jóvenes que no habían practicado ningún sistema de control ni habían recurrido al aborto, resultó ser de seis muertes por 100.000. El estudio mostró, además, que un nivel del ciento por ciento de control de la natalidad y un índice mínimo de mortalidad (menos del 0'5 por 100.000) podría conseguirse en cualquier edad con la combinación de los métodos anticonceptivos tradicionales y el aborto precoz como una ayuda, como un recurso más: verbigracia: utilizando el inocuo pero más inseguro preservativo o diafragma e interrumpiendo cualquier embarazo fortuito mediante el aborto durante el primer trimestre de gestación.

Pese a las decenas de millones de abortos –la mayoría de ellos ilegales– que se practican anualmente, y pese a su repercusión en la salud pública y sobre la fertilidad, el tema ha suscitado escaso interés a nivel internacional. El tema del aborto legal ni siquiera figuró en la agenda de la Conferencia Mundial de la Población de 1974 o en la de la Conferencia Mundial del Año Internacional de la Mujer en 1975; sólo hay una referencia, en el programa de acción que resultó de la última de las convenciones, relativo al aborto ilegal, que se menciona simplemente como uno de los riesgos que contribuyen a elevar la mortalidad materna. Sólo un informe semioficial, limitado a los Estados Unidos, se ocupa del tema de modo claro. *Legalized Abortion and the Public Health*, redactado por el Instituto de Medicina de la Academia Nacional de Ciencias y publicado en 1975, pide que se investiguen "las complicaciones médicas a largo plazo, especialmente después de múltiples abortos, sobre los efectos tanto del aborto como de su prohibición sobre la salud mental y el bienestar social de los individuos y de las familias, y sobre los factores de motivación, conducta y accesos relacionados con la práctica anticonceptiva y la elección de abortar". Los valiosos resultados ya conseguidos por estudios reducidos sobre la legalización del aborto confirman la opinión del Instituto en el sentido de que hace falta más investigación sobre el tema.

Agricultura sin labranza

Dentro de pocos años, muchas de las tierras cultivables de los países desarrollados se plantarán sin vertedera de arado. En la mayoría de las condiciones, plantar sin labranza (pero con herbicidas) puede ahorrar trabajo, energía, agua y suelo

Glover B. Triplett, Jr., y David M. Van Doren, Jr.

Durante siglos, la vertedera del arado ha sido la herramienta básica de la agricultura, que roturaba y removía el suelo en la primera fase de la serie de operaciones que se conoce, en conjunto, como labranza. Sin embargo, en una amplia y creciente área de tierras de labor, el arado empieza a ser desplazado por un sistema agrícola que no involucra la labranza o que reduce grandemente las necesidades de la misma. Las semillas para una nueva cosecha son simplemente plantadas en el suelo que permanece cubierto con los residuos de la cosecha anterior. El control de las malas hierbas, que es el objetivo primario de la labranza, se verifica principalmente mediante la aplicación de herbicidas, y, secundariamente, por el hecho de que el cultivo antiguo actúa como recubrimiento que ahoga el crecimiento de las plantas no deseadas. Las principales ventajas de los nuevos métodos estriban en que reducen el coste del trabajo agrícola y eliminan virtualmente la erosión del suelo por el aire y el agua.

La labranza, tal y como se practicaba hace 40 o 50 años, exigía, para un agricultor que cultivara maíz en el Oeste Medio de Estados Unidos, unas diez pasadas al campo por cosecha. Primero se araba el campo, inmediatamente después se rastrillaba y desmenuzaba los terrones hasta convertirlos en un sustrato fino y firme. Justo antes de que germinara la semilla, se volvía a arar superficialmente para destruir las malas hierbas que habían brotado. Mientras crecía el cultivo, el campo debía ser labrado de dos a cuatro veces más, a fin de controlar las malas hierbas. Estas, si el maíz crecía suficientemente, no se desarrollaban al no llegarles la luz solar; pero si el campo estaba muy infestado, debía labrarse otra vez con una pequeña cultivadora arrastrada por un caballo o un mulo, aun cuando la altura del maíz excediera la de los hombros. Con frecuencia, había que recurrir al trabajo manual para arrancar o cavar con azadones las malas hierbas que sobrevivían a todas las operaciones referidas.

Para el cultivo del algodón, el procedimiento era muy parecido. La producción de granos pequeños, como el trigo, la avena y la cebada, requería menos laboreo, puesto que las plantas crecen tan estrechamente juntas que no existe espacio para la maquinaria de cultivo, una vez han germinado las semillas.

Estos sistemas de labranza se han desarrollado durante un largo periodo de tiempo y, hasta hace poco, se consideraba que representaban el mejor y más digno de confianza de los métodos de cultivo. Preparaban un excelente sustrato y se mostraban muy eficaces en el control de las malas hierbas. Proporcionaban un suelo flojo y aireado que, normalmente, podía absorber bien el agua de la lluvia. La labranza enterraba los restos de las cosechas precedentes, sepultando con ellos ciertos organismos parásitos e insectos que hubieran podido atacar a la nueva cosecha. Por último, la labranza aumentaba la efectividad de los fertilizantes que eran aplicados al suelo.

Y uno se pregunta entonces por qué alguien puede ni siquiera pensar en la eliminación o reducción de estas prácticas tan firmemente arraigadas. La razón escueta es que, bajo condiciones correctas, la eliminación o reducción de la labranza puede hacer más fácil y más provechosa la tarea del agricultor. A pesar de la amplia confianza en el arado, los sistemas que involucran una remoción intensiva del suelo tienen manifiestas desventajas. Son altos los requerimientos que pide la labranza en energía y trabajo, particularmente durante la estación de la siembra. Además, como el periodo óptimo de siembra es bastante corto para muchos cultivos, la cantidad de tierra que los trabajadores de una explotación pueden labrar en el tiempo apropiado es limitada. La labranza elimina de la superficie del suelo los residuos de la cosecha anterior, incrementándose con ello la susceptibilidad del suelo a la erosión, que, a su vez, reduce el potencial agrícola de la tierra. El suelo desnudo se halla expuesto a mayores pérdidas de agua por evaporación y escorrentía que el suelo cubierto por los restos de plantas, con lo que el cultivo dispone de menos agua para su desarrollo.

Uno de los primeros esfuerzos para modificar los métodos de labranza establecidos fue el sistema de arado-sembrado, ideado hace unos 30 años por Ray L. Cook, a la sazón presidente del departamento de ciencias del suelo de la Universidad estatal de Michigan. En este sistema se araba todo el campo, pero las remociones posteriores se limitaban a los surcos sembrados. Estos consistían en hileras de 15 a 25 centímetros de anchura, espaciadas un metro una de otra; se labraban con suma eficacia como para permitir la actuación satisfactoria de la máquina sembradora. El arado inicial dejaba la tierra apta para absorber bien el agua, pero como el suelo entre las líneas se dejaba sin tratar tras el primer labrado, se reducían la erosión y la escorrentía.

El sistema adquirió solamente una aceptación limitada en la zona del maíz. Las condiciones del suelo necesarias para el empleo con éxito del sistema eran

más exigentes que las de la labranza tradicional y cada operación debía ser realizada cuidadosamente para que el sistema funcionara correctamente. Es un axioma de la labranza que las operaciones posteriores corrigen los errores anteriores; pero el sistema de arado-sembrado da pocas oportunidades para su corrección. Además, debía ararse durante la época óptima del sembrado mejor que antes, con lo que la cantidad de tierra que podía manejarse por una fuerza de trabajo de determinada magnitud se veía así fuertemente limitada.

Edward H. Faulkner, que fue agricultor y agente agrícola en Ohio, diseñó uno de los primeros sistemas de labranza que prescindía del arado. En su sistema, el suelo se trabajaba con discos que no profundizaban tanto como la reja del arado ni removían demasiado la tierra. El sistema involucraba también dejar cierta cantidad de residuo de cultivo en el suelo para reducir la erosión. Faulkner describió su método en el libro *Plowman's Folly,* publicado en 1943. Por aquel entonces, arar se consideraba, virtualmente, sinónimo de agricultura, y así el libro daría lugar a muchas disputas.

Una variación del sistema de Faulkner fue la "labranza con recubrimiento". En este sistema, se araba el suelo de alguna manera antes de la siembra, pero luego se aplicaba un recubrimiento de estiércol al tiempo de sembrar o inmediatamente después. El recubrimiento ayudaba a proteger el suelo y a reducir la erosión.

A principios de la década de los cincuenta se desarrolló otro sistema, denominado "labranza-sembrado". Una máquina especial preparaba el suelo en los surcos donde se quería plantar el cultivo y, al mismo tiempo, sembraba las semillas. El espacio que mediaba entre dos surcos se trabajaba por encima con escobillas que limpiaban el suelo sólo lo suficiente para extirpar las malas hierbas.

Los cambios en las prácticas de labranza eran más a menudo el resultado de su aceptación por los agricultores que de los esfuerzos de los investigadores y de los agentes de venta de material agrícola. Cuando se introduce un nuevo concepto o máquina, algunos agricultores evaluarán su rendimiento. Si la producción no mengua y la innovación

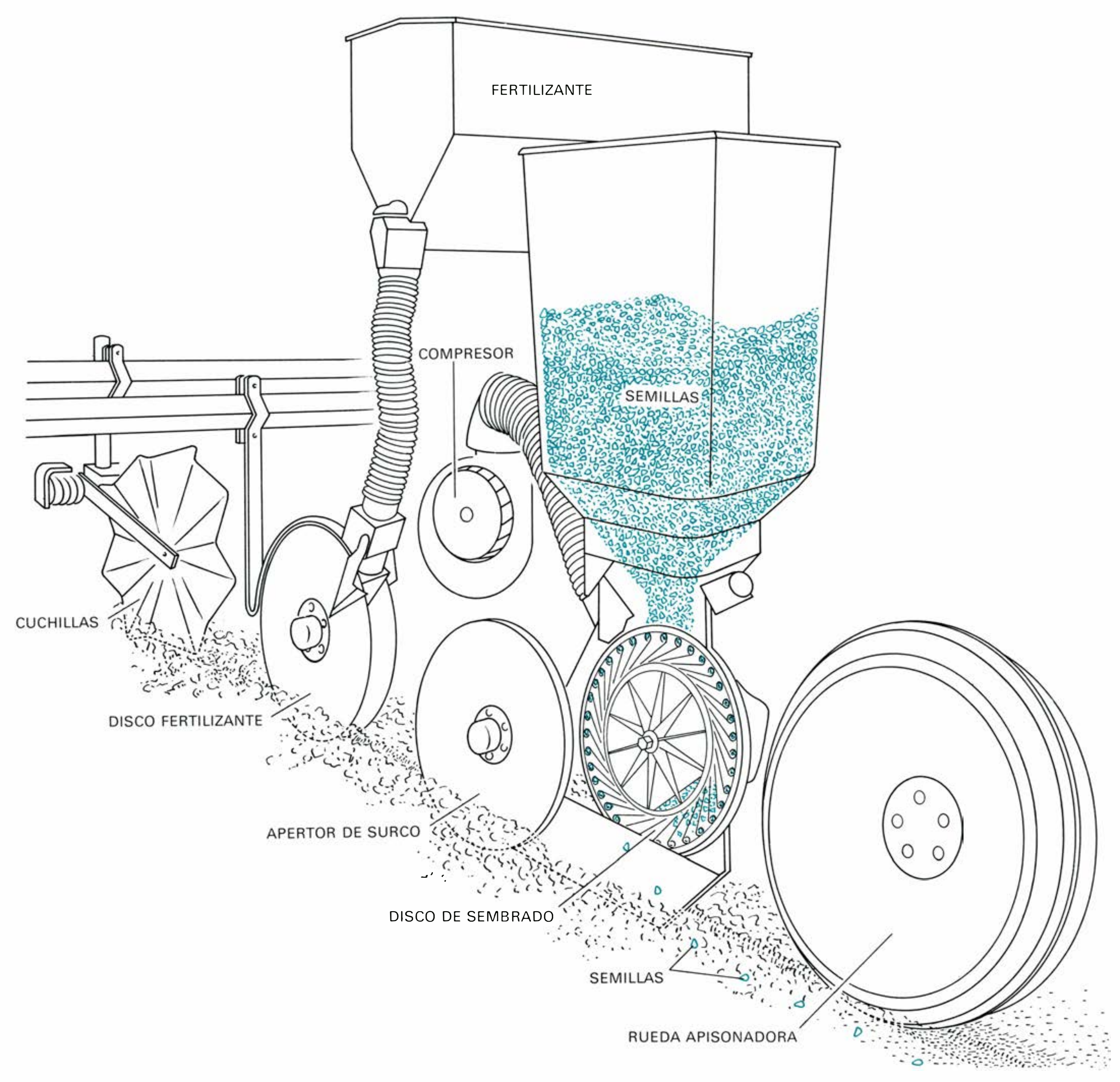


CULTIVO MULTIPLE en una granja de Ohio, Estados Unidos, mediante sembrado sin labranza. La máquina de la derecha recolecta la cosecha de trigo, mientras la máquina de la izquierda siembra semillas de soja entre los rastrojos del mismo. Los dos grandes depósitos blancos de la máquina sembradora contienen abono seco. Delante de ellos, pero no visibles en la fotografía, se encuentran las rejas que abren un surco en el suelo para las semillas de soja, en cuya cavidad se introducen. Las semillas caen de los cuatro depósitos blancos más pequeños. Las cuatro ruedas traseras van cerrando el surco sobre las semillas. Hacia el año 2010, predominará el cultivo que prescinda del arado.

resuelve los problemas o reduce los costes, será aceptada y otros agricultores la adoptarán.

Bajo este criterio, ninguno de los primeros sistemas de labranza que eliminaban el arado tradicional iba demasiado bien. En la fase de desarrollo, los distintos sistemas rendían satisfactoriamente lo suficiente para dar ánimos en lo que se refería a reducir la erosión y a otros problemas asociados con el arado, pero los agricultores que los habían adoptado tenían problemas para controlar las malas hierbas, en particular las malas hierbas perennes, y el volumen de su cosecha disminuía.

La solución al problema de las malas hierbas provino de otro hallazgo, el descubrimiento de los herbicidas selectivos, que tuvo lugar durante la época en la que los tractores reemplazaron a los animales y los fertilizantes nitrogenados se abarataron de tal suerte que su utilización pudo ya generalizarse. El primero de los herbicidas selectivos fue el 2,4-D; fue introducido por los agricultores en las postrimerías de la década de los cuarenta y resultó ser bastante eficaz para el control de las malas hierbas latifolias del maíz. A pesar de que la labranza controla efectivamente las malas hierbas que crecen en el reguero que media entre dos hileras de maíz, las que crecen en las mismas hileras no pueden ser alcanzadas por las máquinas de la-



LA MAQUINA SEMBRADORA empleada en un sistema que prescinde de la labranza realiza varias faenas a la vez. El sentido del movimiento de la máquina, en esta ilustración, es hacia la izquierda. La primera rueda es una cuchilla estriada que abre un reguero estrecho en el suelo no labrado, hundiéndose por entre el mantillo de materia residual de la cosecha precedente, que se deja en la tierra como estiércol. A la cuchilla le sigue un disco que aplica el fertilizante. En este caso se trata de un abono seco, pero en otros se aplica en forma líquida. (El herbicida puede también aplicarse al mismo tiempo que el sembrado para controlar las malas hierbas.) El disco siguiente, separado unos cinco centímetros del anterior abre un surco en donde se echa la semilla. Este disco, a su vez, va seguido por el dispositivo de siembra, que recibe la semilla almacenada en el depósito sito superiormente. Las semillas se deslizan por entre las muescas del dispositivo impulsadas por aire que procede de un compresor; además, el aire mantiene a cada semilla en su lugar hasta que la muesca se acerca al suelo y cae el grano en el surco. La última rueda comprime el suelo sobre la semilla. Varias unidades de este tipo se hallan normalmente juntas en una misma máquina.

branza. Los herbicidas mostraron ser útiles en el control de estas malas hierbas y, por consiguiente, redujeron las faenas manuales.

Los mejores herbicidas selectivos son aquellos compuestos orgánicos que logran sus objetivos a bajas dosis de aplicación y con una baja toxicidad para los animales y el hombre. En las tres décadas transcurridas desde la introducción del 2,4-D se han desarrollado más de 100 herbicidas selectivos para su aplicación en distintos cultivos. Unos herbicidas son eficaces cuando se aplican al follaje de las malas hierbas en crecimiento y otros lo son en el suelo, al inhibir la germinación de las semillas de éstas o al matar las plántulas tras su germinación.

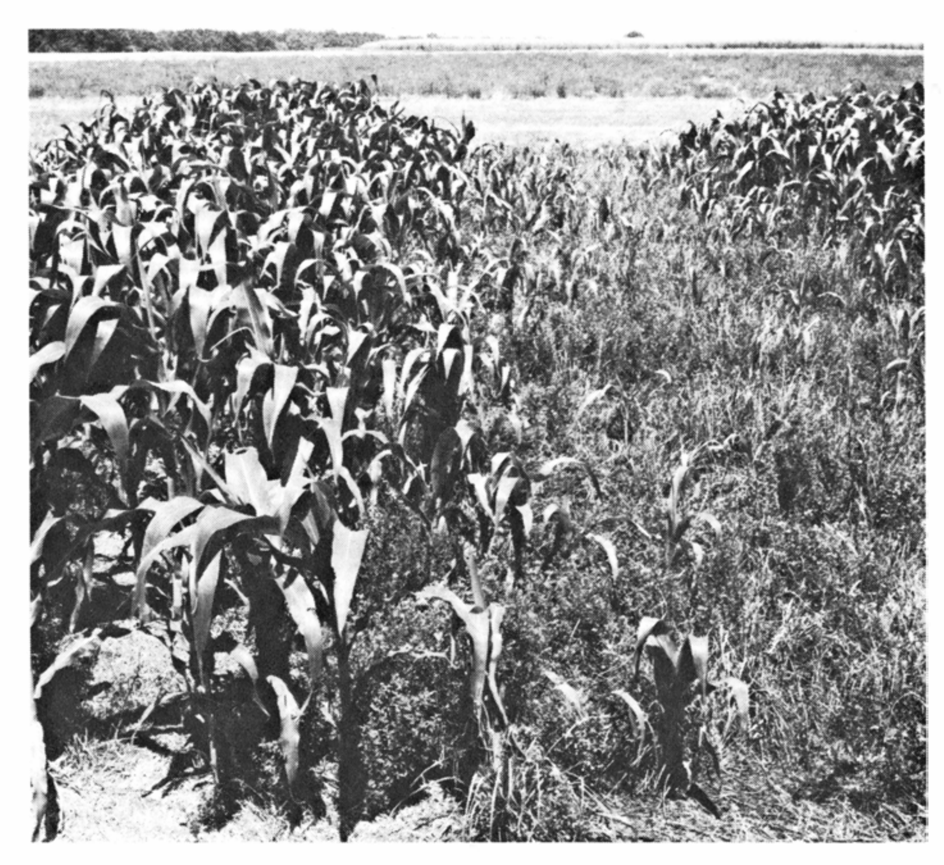
El herbicida exhaustivo, es decir, el herbicida que frene toda la vegetación excepto el sembrado, está todavía por crearse. Muchos herbicidas son eficaces contra grupos de malas hierbas, como hierbas anuales o plantas latifolias. Por consiguiente, se ha convertido en práctica común aplicar combinaciones de dos herbicidas, como mínimo, para ampliar el espectro de las malas hierbas que puedan ser controladas. Para el control de las que se desarrollan en los campos de maíz, ha aprobado la Environmental Protection Agency estadounidense cerca de 20 herbicidas distintos.

Los herbicidas operan multiformemente. Unos inhiben el crecimiento de las plantas infestantes, otros interfieren con la fotosíntesis o con los sistemas enzimáticos y otros terceros marchitan el follaje de las malas hierbas. Es esencial, obviamente, que el herbicida no afecte a la cosecha. Este objetivo puede lograrse gracias a la facilidad del cultivo en metabolizar el herbicida, mediante la aplicación de éste únicamente a las malas hierbas y no al cultivo, o bien mediante la colocación apropiada de las semillas del cultivo respecto de las semillas de las malas hierbas y del herbicida.

Los primeros herbicidas se emplearon como refuerzo de los métodos clásicos de labranza, que continuaron largo tiempo sin modificarse. Con el paso del tiempo y la introducción de herbicidas más eficaces, los agricultores empezaron a eliminar algunas faenas agrícolas previas a la siembra y anteriores a que germinaran las semillas. Redujeron también el número de labores a verificar mientras crecía la cosecha. Los herbicidas controlaban, al mismo tiempo, algunas de las fastidiosas malas hierbas perennes, reduciendo aun más la necesidad de múltiples operaciones de labranza. Con herbicidas eficaces, las técnicas de labranza con recubrimiento, labranza-sembrado y de



CULTIVO DE SOJA que crece por entre un mantillo superficial de maíz residual, en campos experimentales pertenecientes al Ohio Agricultural Research and Development Center. Las raíces, las cañas y las hojas permanecen tal cual las dejó la máquina cosechadora. En esta parcela se han aplicado distintas combinaciones de herbicidas para el cultivo de la derecha y para el de la izquierda. La combinación aplicada al cultivo de la derecha tuvo menos éxito, como lo indican las hierbas que crecen entre la soja.



CULTIVO DE MAIZ, que crece en el mencionado centro Ohio, sobre suelo en el que anteriormente se habían plantado forrajeras. Se aplicaron distintas combinaciones de herbicidas para ahogar el césped y prevenir la aparición de malas hierbas. La combinación central fue la de menor éxito.



LA IMPORTANCIA DEL MANTILLO DEL SUELO, o capa de materia residual dejada en el terreno tras la cosecha del cultivo anterior, en una agricultura sin labranza, se manifiesta en los resultados de las plantaciones experimentales de maíz en un mismo suelo de Ohio (sedimentos margosos de Wooster). La recta negra muestra la producción con labranza tradicional: unos 7205 kilogramos por hectárea. La densidad del mantillo no incide en la producción. La línea de color muestra la producción con sistemas que prescinden de la labranza: 8722 kilogramos por hectárea con un 90 por ciento de mantillo y, 5630 kilogramos por hectárea, con el 5 por ciento.

Faulkner se convirtieron todas ellas en sistemas de trabajo que fueron adoptados ampliamente por los agricultores, y que se continúan empleando en la actualidad, cuando se acentúa la tendencia hacia una menor remoción de la tierra.

La última cota de esta tendencia podría ser un sistema que no involucrara labrar la tierra. Los programas para la mejora de pastos constituyó uno de los primeros pasos en esta dirección. La tierra destinada a pasto suele resultar poco apropiada para la producción de cultivos en surco, porque es escarpada o pedregosa o tiene otras características que hacen la labranza difícil o imposible. Se aplicaron herbicidas a los pastos para destruir la vegetación improductiva, de modo que pudieran establecerse las plantas forrajeras apetecidas.

A nuestro entender, la primera demostración de que un cultivo en hilera puede crecer sin labranza se realizó en Michigan hace unos 20 años, por representantes de una compañía de industrias químicas. Un campo en el cual había crecido césped fue pulverizado con herbicidas para matarlo. Se sembró luego con semillas de trigo por entre el césped muerto. Al mismo tiempo se aplicaron herbicidas repetidamente, con la intención de lograr un buen control de las malas hierbas. A los diez años, cuando se dispuso de herbicidas más eficaces, la práctica de cultivos en hilera sin labranza se adoptó ya, de una manera generalizada, entre los agricultores.

Los sistemas de pulverizado-plantado-recolección actualmente empleados son conocidos por los nombres de "no labranza", "labranza cero", "plantado en césped" y "plantado en surco". Para un cultivo en hilera, como el maíz, el agricultor pulveriza primero con herbicidas el campo, a fin de acabar con toda la vegetación existente y prevenir el desarrollo de malas hierbas a partir de semillas sin germinar en el suelo. La fase siguiente consiste en suministrar abonos, que pueden administrarse en forma seca, si bien lo común es hacerlo en solución, en la que entra también el herbicida. Las semillas se siembran por una máquina que remueve, a través de los restos vegetales, la parte superficial del suelo, echa el grano en el surco y lo recubre; todo ello en una sola operación. El suelo no se altera, a excepción de la franja de cinco a ocho centímetros de anchura abierta por la sembradora. No suelen ser necesarias otras operaciones antes de la época de recolección.

El maíz es la cosecha en surco o hilera más ampliamente cultivada sin labranza. Otros cultivos que se han atendido satisfactoriamente con este método son la soja, el algodón, el cacahuete, el trigo, el tabaco, el sorgo, las forrajeras y algunas hortalizas.

La eliminación de la labranza significa que el agricultor puede confiar completamente a los herbicidas el control de las malas hierbas. Como ningún herbicida podrá realizar esta función con un rendimiento total, habrá que ir tanteando hasta poder decidir con certeza sobre las combinaciones de herbicidas que actuarán mejor. Estas deben ser específicas para el tipo de vegetación preexistente y para el cultivo que se pretenda realizar.

La labranza clásica conseguía mezclar bien el abono con la capa superficial del suelo del campo. Los sistemas sin labranza no permiten estos procesos de mezcla. Pero la ausencia de ese amasijo no constituye ningún problema por lo que se refiere a los tres nutrientes más importantes (nitrógeno, fósforo y potasio) comúnmente aplicados a los cultivos. El nitrógeno es soluble y, por consiguiente, móvil; se introduce hacia el interior del suelo cuando llueve o cuando se riega el campo. El fósforo y el potasio se desplazan con mucho mayor lentitud, pero el recubrimiento por restos, que es una característica de cualquier sistema sin labranza, mantiene la mezcla cerca de la superficie del suelo; así pues, las raíces del cultivo crecen en la zona que contiene inmóviles a los nutrientes.

En el caso de cultivos de forrajeras, se registró una innovación posterior mediante la introducción de otras semillas entre el sembrado del forraje, por ejemplo, semillas de leguminosas en un sembrado de cereales. Hay bastantes leguminosas rentables que no persisten en pastos o praderas y tienen que plantarse periódicamente. Las leguminosas son convenientes como fuente de proteína, para los animales, y, de nitrógeno, para el desarrollo de los cereales.

En muchas zonas, las siembras son de plantas perennes cuyo intervalo de temperaturas óptimas es bastante estrecho, de modo que crecen mejor en las épocas más cálidas del año o en las épocas más frías. En estas siembras se introducen plantas anuales con distintos requerimientos de temperatura, para prolongar el tiempo de producción de la tierra. Por ejemplo, en el sudeste de los Estados Unidos, se siembra trigo de invierno (planta anual de estación fría) en las plantaciones de grama (planta perenne de estación cálida).

Se han dado pasos similares a fin de sacar mayor provecho de la estación de crecimiento en el caso de cosechas anuales. Se ha ensayado, por ejemplo, el establecimiento de un segundo cultivo después de la recolección del primero o antes de que la cosecha llegue a su maduración final. Una disposición de cosecha múltiple, común en los Estados Unidos, consiste en plantar soja tras la recolección de una cosecha de grano pequeño. Los métodos de sembrado sin labranza han contribuido significativamente a esta práctica.

La cosecha de grano pequeño se recolecta algunas semanas después de la época óptima para la siembra de la soja. Si ha de sucederle realmente el cultivo de soja, es necesario sembrarla en seguida; germinará y se desarrollará muy

pronto. Las máquinas sembradoras están dotadas de rejas de arado o de muelas que levantan la tierra por entre los rastrojos del cultivo de grano pequeño anterior, de forma que las semillas de soja puedan ser plantadas inmediatamente después de la recolección del grano. Se aplican herbicidas de pregerminación para matar las malas hierbas que se encuentran entre los rastrojos del grano pequeño y para prevenir la aparición de malas hierbas, una vez plantada la soja. Los límites septentrionales del cultivo doble, en los Estados Unidos, se han extendido varios centenares de kilómetros merced a los métodos de plantado sin labranza.

Hemos dicho que la reducción de la erosión es el beneficio más importante que se consigue con la reducción o eliminación de la labranza. La reducción es significativa para la mayoría de los casos en que se combine menos labranza con un recubrimiento del suelo por los rastrojos; si bien las reducciones más drásticas de la erosión proceden de los sistemas que eliminan la labranza. En 1969, se hizo una prueba de campo en la North Appalachian Experimental Watershed Research Station, perteneciente al Departamento de Agricultura, próxima a Coshocton, Ohio. Se plantó maíz en condiciones de no labranza y de labranza tradicional, en algunos declives cuya extensión superficial oscilaba entre media y algunas hectáreas. Con instrumentos apropiados se medía la pérdida de agua y de suelo de los campos durante las tormentas. Una lluvia de 125 mm, caída durante el mes de julio, época en la cual los campos labrados al modo tradicional son particularmente vulnerables a la erosión, causó pérdidas de más de 40 toneladas de suelo por hectárea, en tales campos, con pendientes del 6 al 8 por ciento. Las pérdidas de suelo en declives con una pendiente del 20 por ciento, donde el maíz había sido plantado sin labranza, en rastrojo muerto, no llegó a los 90 kilogramos por hectárea.

Una pendiente del 20 por ciento resulta demasiado escarpada para cultivos en surco con labranza tradicional, a causa de la destrucción por erosión. Una pendiente tal se halla también muy cerca del límite de seguridad para las faenas del equipo agrícola. Como la erosión puede reducirse a la centésima parte o más, mediante el sembrado sin labranza, tenemos pues que se hace viable con este sistema el cultivo en hilera en terrenos ondulantes. Los suelos más altamente productivos suelen hallarse en áreas montañosas; por ello, la práctica los ha relegado al cultivo de forrajeras como medida de conservación. Con los métodos de no labranza, una elevada proporción de esta tierra podrá sembrarse de plantas más provechosas.

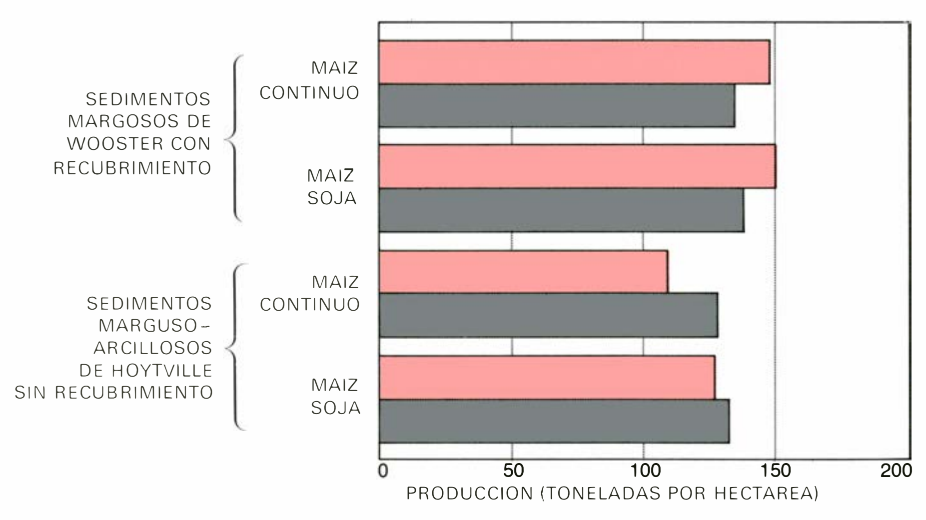
La escorrentía de los campos en que la labranza es mínima determina una marcada disminución de la carga de sedimentos, de forma que queda mejorada la calidad del flujo. Como los herbicidas van esencialmente asociados a los métodos que reducen o eliminan la labranza, hemos controlado la escorrentía de los herbicidas químicos. El transporte de herbicidas disueltos en agua de escorrentía es más elevado cuando la lluvia cae dentro de las horas inmediatas a la aplicación del herbicida, pero el movimiento total de herbicidas resulta casi despreciable. Lo hemos medido y resulta ser de pocos gramos por hectárea para lluvias que aparecen dentro de los dos días después de la aplicación del herbicida. Sin embargo, cuando la lluvia es lo suficientemente violenta como para producir una cantidad significativa de escorrentía, se produce cierto movimiento de herbicidas. En una ladera, cuya extensión superficial alcanzaba varios miles de hectáreas, se sembró y se aplicaron herbicidas en distintos campos, durante un periodo de cuatro a seis semanas. Así, de producirse lluvias suficientemente enérgicas como para causar un flujo significativo de herbicidas, afectarían solamente a los campos más sensibles a la pérdida de éstos.

Otra ventaja de los sistemas que hemos ido describiendo es su economía de tiempo y trabajo. Los métodos de la labranza clásica requieren una cantidad considerable de energía y trabajo para la preparación del sustrato. Aunque, en algunas situaciones, parte de la labranza puede hacerse antes de la estación de siembra, la práctica no es aplicable en campos en los que la erosión es un riesgo. Un problema que se añade a los que tienen que hacer frente los agricultores que disponen solamente de un espacio de tiempo corto para labrar, lo constituyen los días lluviosos que obligan a suspender las faenas agrícolas. En Estados Unidos, los agricultores que viven en la franja que se extiende desde el este de Iowa hasta la costa atlántica tienen que afrontar los riesgos de lluvia y el suelo mojado en la época de siembra.

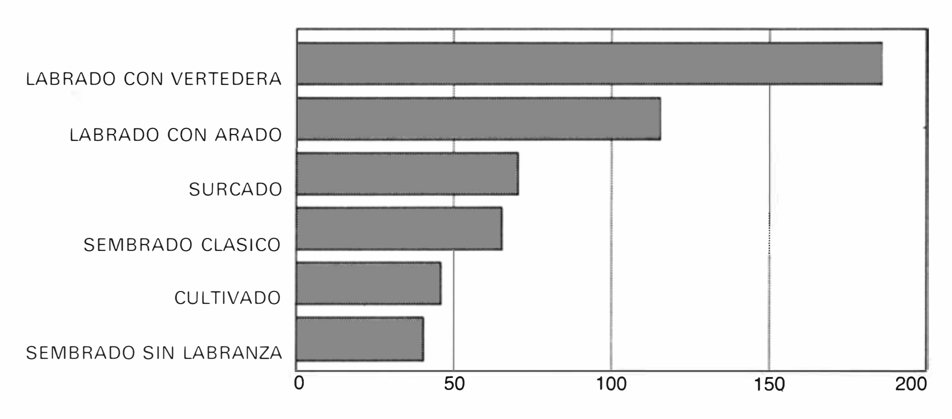
El sembrado directo, en suelo no labrado, con equipo adecuado, es una faena rápida, en la que se gasta relativamente poca energía, ya que se reducen las necesidades de los grandes tractores. Siempre que la humedad del suelo sea favorable para la labranza de cualquier tipo, puede ponerse en acción máquinas sembradoras.

Otra faena esencial para la siembra sin labranza es la aplicación de herbicidas, que puede realizarse también en seguida, a veces conforme se va sembrando. En general, de la adopción de métodos que prescinden de la labranza puede esperarse un incremento en un tercio, aproximadamente, de la productividad por parte de los trabajadores agrícolas. La cantidad de combustible necesario para el establecimiento de la cosecha puede reducirse unos dos tercios de la que se consume en la labranza clásica.

Una vez determinadas las exigencias necesarias para el control de las ma-



LA IMPORTANCIA DEL TIPO DE SUELO, junto con las condiciones de recubrimiento, en la agricultura sin labranza, se muestra en los resultados obtenidos de plantaciones experimentales en dos suelos distintos de Ohio: el sedimento margoso de Wooster, perfectamente drenado, y el sedimento margusoarcilloso de Hoytville, peor drenado. Tanto para el cultivo continuo de maíz como para el cultivo alterno de maíz y soja, la producción en el sedimento margoso de Wooster fue mayor sin labranza (*barras coloreadas*) que con el sistema de labranza tradicional (*barras grises*).



COMPARACION DE COSTES de la energía que se necesita en las faenas preparatorias de la agricultura sin labranza y de la labranza tradicional, tomando como punto de referencia el combustible que gasta un tractor diesel de 100 caballos. Los costes de energía de la labranza tradicional comprenden no sólo los costes del arado, sino también los del surcado, siembra y cultivo. Los datos para este diagrama y para los restantes de estas páginas fueron recogidos por los autores y sus colegas del Ohio Agricultural Research and Development Center, Estados Unidos.

las hierbas y el establecimiento del cultivo, el factor más importante en la definición de la respuesta de un cultivo a la labranza es el drenaje del suelo. A mejor drenaje, menos labranza ne-
En suelos bien drenados, el sis-
cesaria.
tema de no labranza se revela como el método idóneo para la potenciación de la cosecha, siempre que, como mínimo, el 60 por ciento de la superficie esté cubierta por la materia residual de la cosecha anterior. Para ponderar en su justo valor esto, piénsese en que los suelos mejor drenados se encuentran, a menudo, en los terrenos pedregosos, en donde la erosión es un problema.

El mantillo residual de la cosecha precedente conserva la humedad, no sólo al reducir la evaporación sino también al incrementarse la captura de agua de lluvia, que tiende a perderse cuando los suelos están desnudos. La materia residual mantiene, además, la profundidad del suelo, o estructura útil, que se preserva al no destruirse su estructura por la labranza. Además, las lombrices de tierra suelen abundar bajo el mantillo de materia residual del campo no labrado, y, cuanto menos son molestadas, más ayudan en el desarrollo de una estructura apropiada del suelo. En estas condiciones, el maíz responde favorablemente produciendo de un 10 a un 20 por ciento más que el maíz que crece en suelos de labranza tradicional. Durante varios años hemos medido la cantidad de producción en campos de suelos bien drenados, en los que se ha plantado maíz continuamente y otros campos en los que el maíz se planta cada tres años en rotación con otros dos cultivos. Los campos plantados sin labranza han superado la producción de los labrados al uso tradicional en unos 1400 kilogramos por hectárea y año, donde el maíz se cultivó sin solución de continuidad, y, en unos 550 kilogramos, donde se plantó maíz según un sistema que era de rotación trianual.

Los métodos que prescinden de la labranza no pueden llevarse a la práctica en cualquier tipo de suelo o condición edáfica. En los suelos pobremente drenados, el mantillo de materia residual retiene más humedad en superficie al principio de la primavera que lo que sería de desear. La humedad, a su vez, retrasa el calentamiento del suelo y la tasa a la cual el cultivo germina y crece. Por otro lado, una temperatura inferior bajo el mantillo de residuos puede constituir una ventaja en zonas tropicales, en donde la temperatura de la superficie del suelo desnudo puede alcanzar los 50° C (122° Fahrenheit), que es lo suficientemente alta para dañar a las plántulas.

Hemos comprobado que, en los suelos escasamente drenados, la producción de maíz, en régimen de cultivo continuo, es cerca de un 30 por ciento menor con los métodos que prescinden de la labranza que con aquellos otros en que se ara el suelo. Estas diferencias tardan dos o tres años en manifestarse, y no aparecen cuando al maíz sucede algún otro cultivo.

Es de sospechar que se desarrollen organismos patógenos en suelos pobremente drenados, si éstos no se labran. Las producciones son equivalentes para maíz, en régimen de cultivo continuo, con labranza clásica y sin labranza, si se aró la tierra el otoño o el invierno precedentes.

Nada más lejos de nosotros que dar a entender que los sistemas de labranza reducida no tienen problemas específicos. Un agricultor que cambie a un sistema tal deberá conocer el fracaso antes de adquirir ciertas prácticas, si pretende explotar el sistema con éxito. Estas prácticas se refieren a los métodos de plantación, al uso de productos químicos y al método de roturar la tierra que sea apropiado dadas las condiciones del suelo.

Plagas, malas hierbas e insectos pueden constituir un problema más importante con los sistemas de labranza reducida que siguiendo el sistema de labranza tradicional. A medida que se reducen las faenas de labranza propiamente dicha, se incrementa la dependencia del agricultor respecto del empleo adecuado de herbicidas. La labranza favorece el control de insectos, toda vez que destruye sus hábitats, en tanto que un mantillo de materia residual proporciona cobertura y nichos ideales para el depósito de huevos de diversos insectos nocivos. Las orugas, que constituyen un problema incidental en los campos labrados, han destruido extensiones enteras de maíz plantadas, prescindiendo del arado, en césped muerto.

Las babosas se multiplican en campos cubiertos con residuos. La población de ratones suele ser más alta en campos tapizados por un mantillo de gran espesor; los ratones se comen las semillas del maíz antes de que la planta enraíce. Todas estas plagas pueden controlarse con éxito, pero el agricultor debe estar atento, cuidar la identificación de tales problemas e intervenir antes de que sean insolubles.

Tomando como punto de referencia los sistemas de labranza seguidos hace 40 o 50 años, nos encontramos con que se ha registrado un fuerte movimiento hacia la adopción de sistemas de labranza reducida. Un estudio realizado recientemente por el Soil Conservation Service indica que en los Estados Unidos, en 1976, se han plantado, sin labranza, casi 3,2 millones de hectáreas, y que otros 23,3 millones de hectáreas adicionales han reducido el nivel que tradicionalmente poseían. La cantidad de tierras labradas según sistemas clásicos ha alcanzado los 100 millones de hectáreas aproximadamente. En un informe sobre labranza mínima publicado en 1975, el Departamento de Agricultura estadounidense pronosticaba que, en el año 2010, más del 90 por ciento de los cultivos crecerían con sistemas de labranza reducida y que, en más de la mitad de los mismos, se practicaría alguna forma de agricultura sin labranza.

La mecánica cuántica de los agujeros negros

Los agujeros negros suelen definirse como regiones de las que nada, ni siquiera la luz, puede escapar. Sin embargo, hay buenas razones para creer que las partículas sí salen, en virtud del "efecto túnel"

S. W. Hawking

Los primeros 30 años de este siglo vieron el nacimiento de tres teorías que han cambiado radicalmente la opinión de los hombres sobre la física y la realidad misma. Los físicos están tratando todavía de explorar sus implicaciones y de armonizarlas conjuntamente. Las tres teorías fueron la teoría de la relatividad especial (1905), la teoría de la relatividad general (1915) y la teoría de la mecánica cuántica (c. 1926). A Albert Einstein se debe gran parte de la primera, se le debe toda la segunda y desempeñó un importante papel en el desarrollo de la tercera. Sin embargo, Einstein no aceptó nunca la mecánica cuántica al integrar ésta, entre sus componentes, la casualidad y la incertidumbre. Su pensamiento se resume en la frase repetidamente citada: "Dios no juega a los dados". No obstante, la mayoría de los físicos aceptaron rápidamente la relatividad especial y la mecánica cuántica porque describían efectos que podían observarse de una manera directa. La relatividad general, por el contrario, apenas si mereció la más mínima atención, porque parecía demasiado complicada desde el punto de vista matemá.ico, no podía someterse a comprobación en el laboratorio y era una teoría puramente clásica que no parecía compatible con la mecánica cuántica. De esta manera, la relatividad general permaneció en la penumbra durante casi 50 años.

La proliferación de observaciones astronómicas, que comenzó a principios de los años 60, reavivó el interés por la teoría clásica de la relatividad general, porque parecía que muchos de los nuevos fenómenos que se estaban descubriendo, como los quasars, pulsars y fuentes densas de rayos X, apuntaban hacia la existencia de campos gravitatorios muy fuertes, campos que únicamente podían ser descritos por la relatividad general. Los quasars son objetos semejantes a estrellas, cuyo brillo debe ser muchas veces mayor que el de galaxias enteras si se hallan tan alejados como indica el corrimiento hacia el rojo de sus espectros; los pulsars son restos, de rápida pulsación, de explosiones de supernovas, y se cree que son estrellas ultradensas de neutrones; las fuentes densas de rayos X, reveladas por instrumentos a bordo de vehículos espaciales, pueden ser también estrellas de neutrones u objetos hipotéticos de densidad todavía mayor, llamados agujeros negros.

Uno de los problemas que se presentaron a los físicos que intentaban aplicar la relatividad general a estos descubrimientos nuevos u objetos hipotéticos fue hacerla compatible con la mecánica cuántica. En los últimos años se han registrado progresos que permiten confiar en que pronto tendremos una teoría cuántica de la gravitación totalmente carente de contradicciones internas, que estará de acuerdo con la relatividad general para los objetos macroscópicos y, asimismo, libre de los infinitos matemáticos que han plagado por mucho tiempo otras teorías cuánticas de campos. Estos avances tienen que ver con ciertos efectos cuánticos, recientemente descubiertos, asociados con los agujeros negros, los cuales señalan una notable relación entre los agujeros negros y las leyes de la termodinámica.

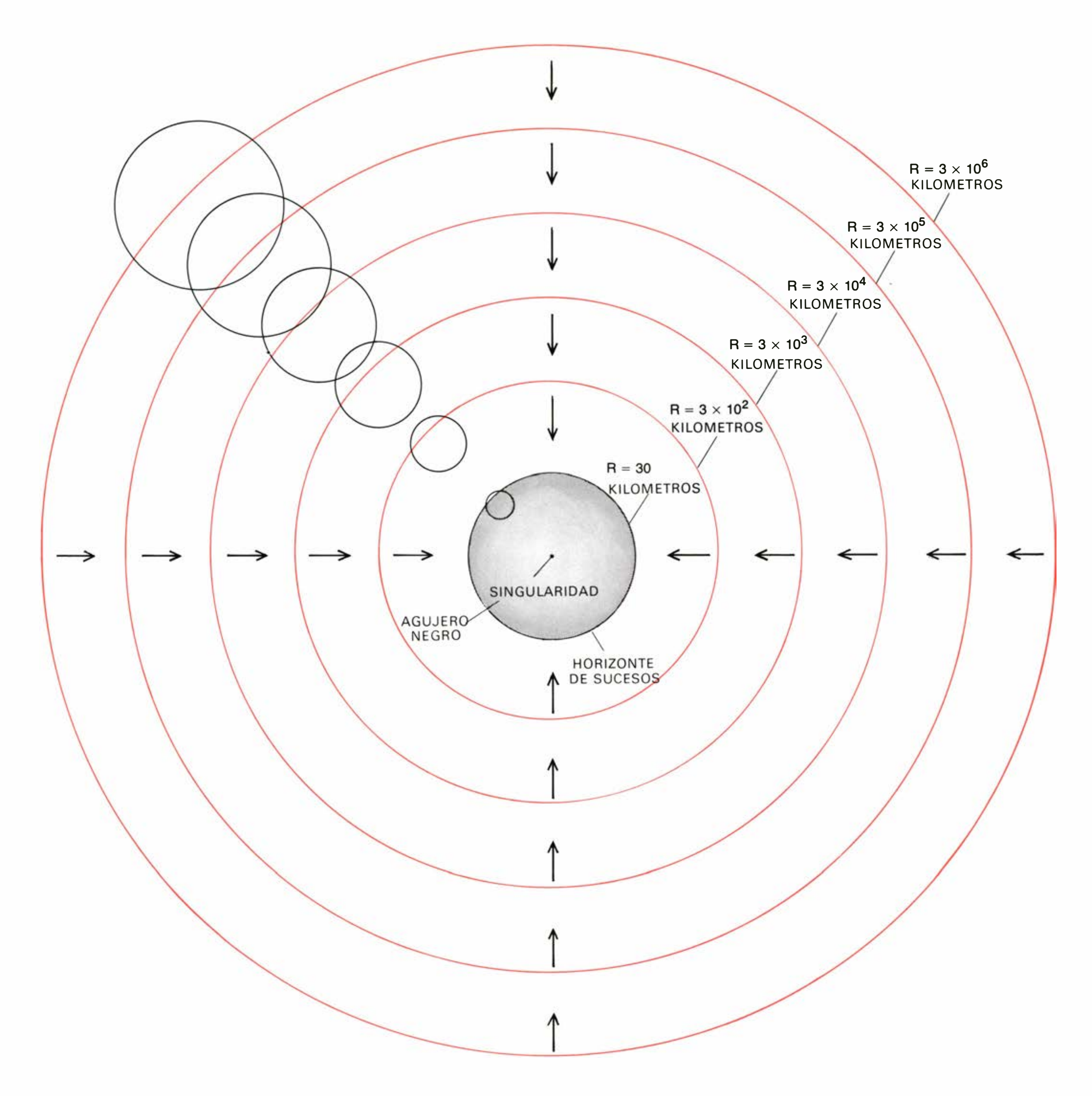
Déjeseme describir brevemente como se puede crear un agujero negro. Imaginemos una estrella con una masa 10 veces mayor que la del sol Durante la mayor parte de su vida, que dura mil millones de años aproximadamente, la estrella generará calor en su centro al convertir hidrógeno en helio. La energía liberada creará suficiente presión para mantener a la estrella contra su propia gravedad, originando un objeto cuyo radio será unas cinco veces mayor que el radio del sol. La velocidad de escape, desde la superficie de tal estrella, sería de unos 1000 kilómetros por segundo. Esto significa que un objeto lanzado verticalmente hacia arriba desde la superficie de la estrella, con una velocidad inferior a 1000 kilómetros por segundo, sería atrapado por el campo gravitatorio de la estrella, y volvería a la superficie; por el contrario, un objeto dotado de una velocidad mayor escaparía al infinito.

Cuando la estrella hubiera acabado su combustible nuclear, no habría nada que mantuviera la presión hacia fuera, y la estrella empezaría a colapsarse en razón de su propia gravedad. Al disminuir el tamaño de la estrella, el campo gravitatorio de su superficie sería cada vez más fuerte y aumentaría la velocidad de escape. En el momento en que el radio se hubiera reducido hasta 30 kilómetros, la velocidad de escape hubiera crecido hasta 300.000 kilómetros por segundo, es decir, la velocidad de la luz. A partir de ese momento, la luz emitida por la estrella no podría escapar al infinito sino que quedaría atrapada por el campo gravitatorio. De acuerdo con la teoría de la relatividad especial, nada puede viajar más rápido que la luz, de forma que, si la luz no puede escapar, no puede tampoco escapar nada más.

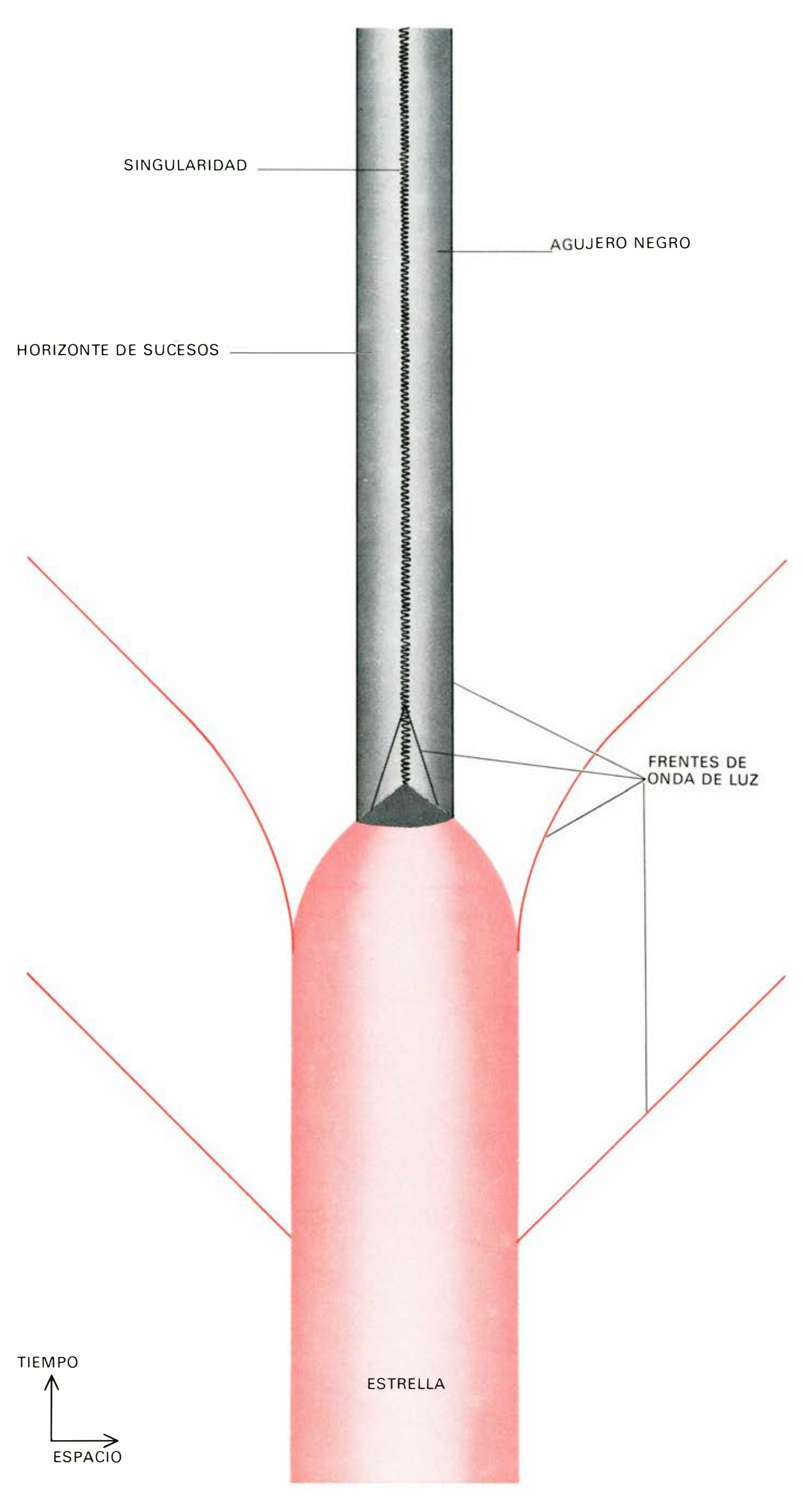
El resultado sería un agujero negro; o sea, una región del espacio-tiempo desde la que no se puede huir al infinito. El contorno del agujero negro se llama el

horizonte de sucesos. Corresponde a un frente de onda de luz procedente de la estrella que no puede escapar al infinito sino que va manteniéndose en el radio de Schwarzschild: $2GM/c^2$, donde G es la constante de la gravitación de Newton, M es la masa de la estrella y c la velocidad de la luz. Para una estrella de unas 10 masas solares, el radio de Schwarzschild alcanza los 30 kilómetros.

Disponemos hoy de buenas pruebas experimentales que postulan la existencia de agujeros negros de este tamaño en sistemas de estrellas dobles, tales como la fuente de rayos X conocida como Cygnus X-1. También podría haber bastantes agujeros negros, mucho más pequeños, esparcidos por el universo, formados no por el colapso de estrellas sino por el colapso de regiones fuertemente comprimidas en el medio caliente y denso que se cree existió poco después de la gran explosión en la que se



COLAPSO DE UNA ESTRELLA de 10 masas solares representado esquemáticamente. Se parte aquí de un radio original de tres millones de kilómetros (unas cinco veces el radio del sol) hasta llegar a los 30, cuando ella desaparece dentro del "horizonte de sucesos", que define el límite exterior de un agujero negro. La estrella continúa colapsándose hacia lo que se llama una singularidad espaciotemporal, sobre la que nada dicen las leyes de la física. La serie de seis pequeños círculos representa los frentes de onda de la luz emitida desde las superficies sucesivas, un instante antes de que la estrella se haya colapsado hasta las dimensiones mostradas. Los radios de la estrella y de los frentes de onda están en escala logarítmica. En el transcurso del colapso, va cayendo más frente de onda dentro del volumen de la estrella a medida que la velocidad de escape aumenta desde 1000 kilómetros por segundo hasta 300.000 kilómetros por segundo, que es la velocidad de la luz. La velocidad final se alcanza cuando la estrella llegue a desaparecer en el interior del horizonte de sucesos.



EL COLAPSO GRAVITATORIO DE UNA ESTRELLA se representa en este diagrama espacio-tiempo, en el que se han suprimido dos de las tres dimensiones espaciales. La dimensión vertical es el tiempo. Cuando el radio de la estrella alcanza un valor crítico, el radio de Schwarzschild, la luz emitida por la estrella ya no puede escapar, sino que se mantiene en este radio, formando el horizonte de sucesos, o contorno del agujero negro. Dentro del agujero negro la estrella continúa colapsándose hasta una singularidad.

originó el universo. Tales agujeros negros "primarios" son del mayor interés por los efectos cuánticos que describiré a continuación. Un agujero negro que pesara mil millones de toneladas (aproximadamente la masa de una montaña), tendría un radio de unos 10^{-13} centímetros (el tamaño de un neutrón o un protón). Podría estar girando en órbita alrededor del sol o del centro de la galaxia.

El primer indicio de que podría haber una conexión entre los agujeros negros y la termodinámica se produjo con el descubrimiento matemático, en 1970, de que el área de la superficie del horizonte de sucesos, el contorno de un agujero negro, siempre aumenta cuando cae materia o radiación adicional en el agujero negro.

Además, si chocan dos agujeros negros y se unen formando otro único, el área del horizonte de sucesos alrededor de este es mayor que la suma de las áreas de los horizontes de sucesos de los agujeros negros originales. Estas propiedades manifiestan que hay una similitud entre el área del horizonte de sucesos de un agujero negro y el concepto de entropía de la termodinámica. La entropía se puede considerar como la medida del desorden de un sistema o, análogamente, como un desconocimiento acerca de cuál es precisamente su estado. La famosa segunda ley de la termodinámica dice que la entropía siempre aumenta con el tiempo.

La analogía entre las propiedades de los agujeros negros y las leyes de la termodinámica ha sido analizada por James M. Bardeen, de la Universidad de Washington, por Brandon Carter, que trabaja en el Observatorio de Meudon, y por el autor. La primera ley de la termodinámica dice que un pequeño cambio en la entropía de un sistema va acompañado de un cambio proporcional en la energía del sistema. El factor de proporcionalidad se llama temperatura del sistema. Bardeen, Carter y el autor encontraron una ley similar que relaciona el cambio en la masa de un agujero negro con un cambio en el área del horizonte de sucesos. Aquí el factor de proporcionalidad comprende una cantidad llamada gravitación superficial, que mide la intensidad del campo gravitatorio en el horizonte de sucesos. Si se acepta que el área del horizonte de sucesos es análoga a la entropía, entonces parecería que la gravitación superficial es análoga a la temperatura. La analogía aumenta por el hecho de que la gravitación superficial es la misma en todos los puntos del horizonte de sucesos, de igual forma que la

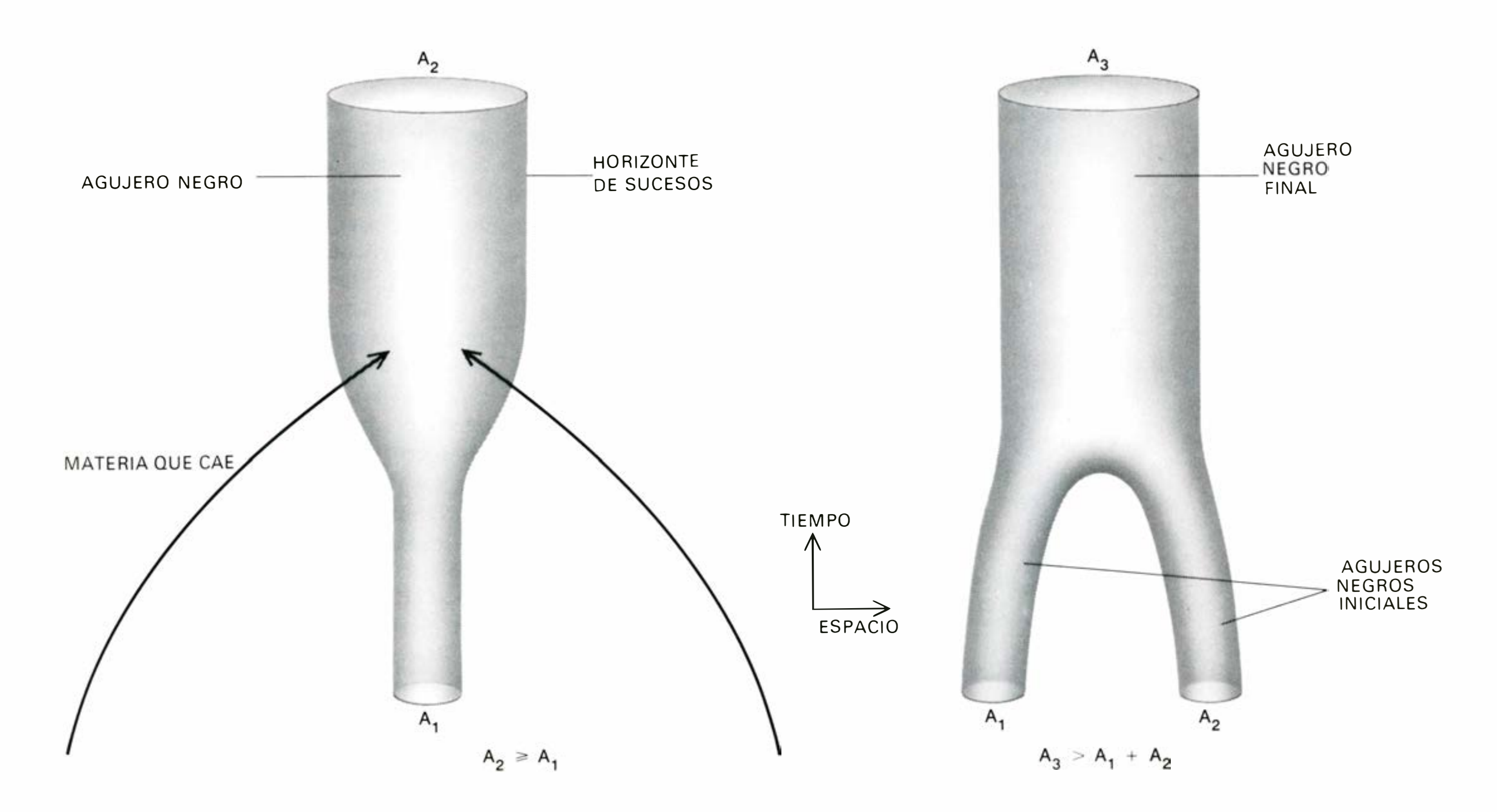
temperatura es la misma en cualquier punto de un cuerpo que esté en equilibrio térmico.

Aunque había una clara similitud entre la entropía y el área del horizonte de sucesos, no parecía nada obvio que el área pudiera asimilarse a la entropía de un agujero negro. ¿Qué se entendía por entropía de un agujero negro? La sugerencia crucial fue hecha, en 1972, por Jacob D. Bekenstein, que era entonces estudiante graduado en la Universidad de Princeton y profesa ahora en la del Negev, en Israel. Es la siguiente: cuando un agujero negro se crea por colapso gravitatorio, pasa rápidamente a un estado estacionario caracterizado únicamente por tres parámetros, a saber, la masa, el momento angular y la carga eléctrica. Aparte de estas tres propiedades, el agujero negro no conserva ningún otro rasgo del objeto que se colapsó. Esta conclusión, conocida como el teorema de "un agujero negro no tiene pelo" fue demostrada por el trabajo combinado de Carter, Werner Israel, de la Universidad de Alberta, David C. Robinson, del King's College de Londres y el autor.

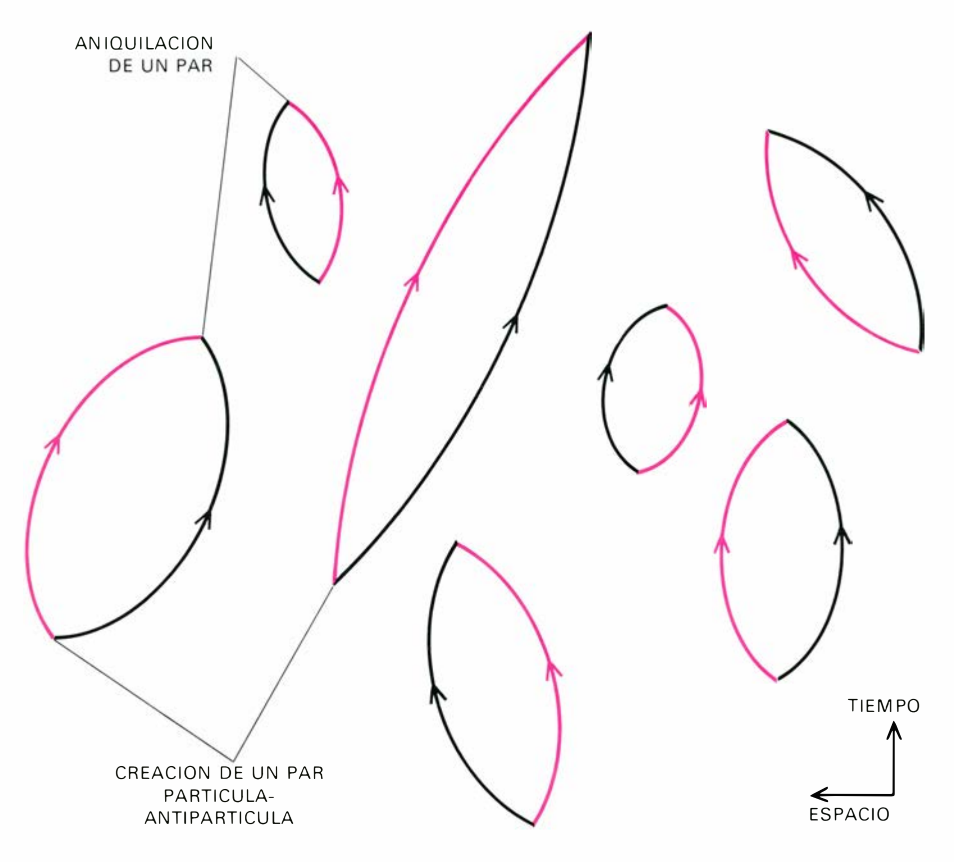
El teorema anterior implica que se pierde una gran cantidad de información en el colapso gravitatorio. Por ejemplo, el estado final del agujero negro es independiente de si el cuerpo que se colapsó estaba compuesto de materia o de antimateria, y de si era esférico o de forma muy irregular. En otras palabras, un agujero negro de una masa, momento angular y carga eléctrica determinados podría haber sido formado por el colapso de cualquiera de las numerosísimas configuraciones diferentes de materia. Aún más, si se desprecian los efectos cuánticos, el número de configuraciones sería infinito, ya que el agujero negro podría haberse formado por el colapso de una nube de un gran número de partículas de masa ilimitadamente pequeña.

El principio de incertidumbre de la mecánica cuántica, implica, sin embargo, que una partícula de masa m se comporte como una onda de longitud de onda h/mc; donde h es la constante de Planck (una cifra bajísima: $6{,}62 \times 10^{-27}$ erg-segundo) y c es la velocidad de la luz. Para que una nube de partículas se colapse y forme un agujero negro, pudiera parecer necesario que esta longitud de onda fuera menor que el agujero negro que se constituya. Por ello, aunque el número de configuraciones que podrían formar un agujero negro de una masa, momento angular y carga eléctrica dadas es muy grande, debe ser finito. Bekenstein sugirió que el logaritmo de este número podría interpretarse como la entropía de un agujero negro. El logaritmo del número sería una medida de la cantidad de información irremediablemente perdida durante el colapso a través del horizonte de sucesos cuando se creaba un agujero negro.

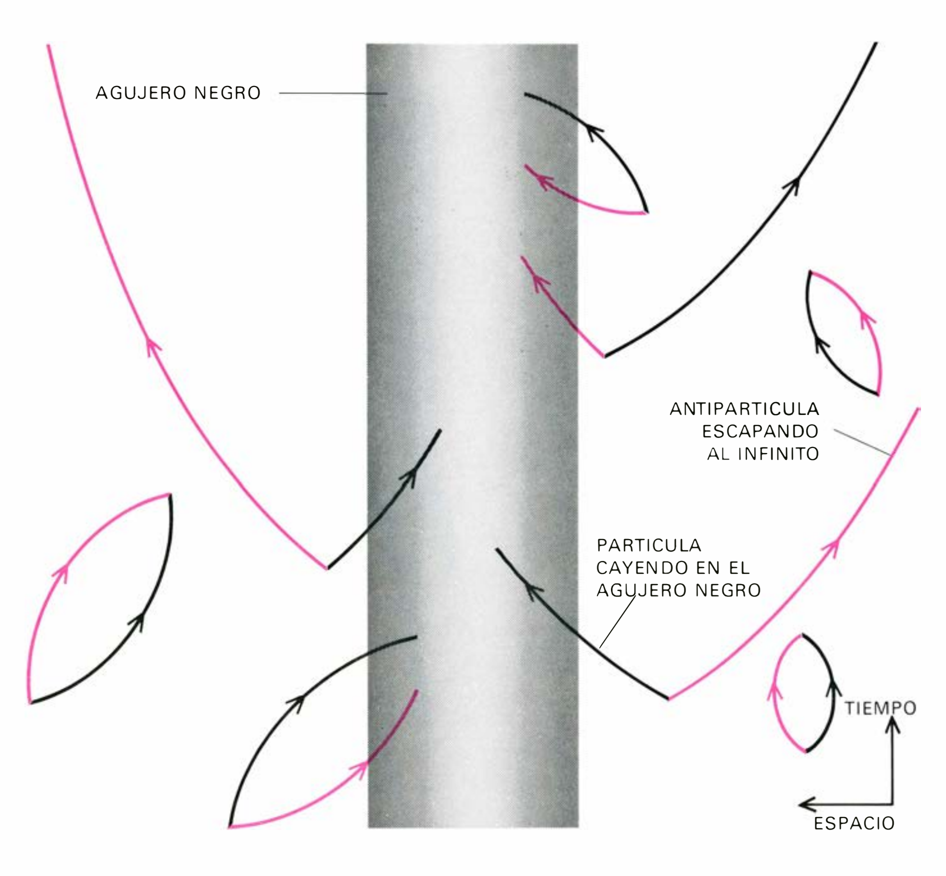
El error, aparentemente fatal, de la sugerencia de Bekenstein era que, si un agujero negro tenía una entropía finita, proporcional al área de su horizonte de sucesos, debería tener también una temperatura finita, que sería proporcional a su gravitación superficial. Esto implicaría que un agujero negro podría estar en equilibrio con la radiación térmica a temperatura distinta de cero. Pero, de acuerdo con la teoría clásica, un equilibrio tal no es posible, ya que el agujero negro absorbería cualquier radiación térmica que cayera en su interior; ahora bien, por definición, no sería capaz de emitir ninguna.



ALGUNAS PROPIEDADES de los agujeros negros sugieren que hay una similitud entre el área del horizonte de sucesos de un agujero negro y el concepto de entropía de la termodinámica. A medida que la materia y la radiación van cayendo en el agujero negro *(configuración espacio-tiempo de la izquierda)*, **el área de la sección transversal del horizonte de sucesos aumenta paulatinamente. Si dos agujeros negros chocan y se unen** *(configuración de la derecha)*, **el área de la sección transversal del horizonte de sucesos del agujero negro ocasionada por el choque es mayor que la suma de las áreas de los horizontes de sucesos pertenecientes a los dos agujeros negros que se unieron entre sí.**



EL ESPACIO-TIEMPO "VACIO" está lleno de pares "virtuales" de partículas *(negro)* y antipartículas *(color)*. Los dos componentes de un par se crean simultáneamente en un punto del espacio-tiempo, se separan y se juntan de nuevo aniquilándose mutuamente. Se llaman virtuales porque, a diferencia de lo que ocurre con las partículas "reales", no pueden detectarse.



EN LAS CERCANIAS DE UN AGUJERO NEGRO un componente de un par partícula-antipartícula puede caer en el agujero, dejando al otro componente sin compañero con el que aniquilarse. Si el superviviente del par no sigue a su compañero hacia el agujero negro, puede escapar al infinito. De esta forma, el agujero negro parece emitir partículas y antipartículas.

Esta paradoja se mantuvo hasta principios de 1974, cuando el autor se hallaba investigando el comportamiento de la materia en las proximidades de un agujero negro, según la mecánica cuántica. Para gran sorpresa suya encontró que el agujero negro parecía emitir partículas de una manera uniforme. Según era opinión común entonces, aceptó que un agujero negro no podía emitir nada. Por esta razón, puso gran empeño en tratar de eliminar este embarazoso efecto. No hubo manera de hacerlo, por lo que al final tuvo que aceptarlo. Lo que finalmente le convenció de que era un proceso físico real fue que las partículas salientes tenían un espectro cabalmente térmico: el agujero negro crea y emite partículas y radiaciones como si fuera un cuerpo caliente ordinario, con una temperatura que es proporcional a la gravitación superficial e inversamente proporcional a la masa. Esto hizo que la hipótesis de Bekenstein, sobre la entropía finita del agujero negro, resultara totalmente coherente, ya que implicaba que un agujero negro podía estar en equilibrio térmico a temperaturas finitas distintas de cero.

Desde entonces, las pruebas matemáticas de que un agujero negro puede emitir térmicamente han sido confirmadas por otros siguiendo distintos enfoques. Una manera de entender la emisión es la siguiente: la mecánica cuántica presupone que el espacio está lleno de pares de partículas y antipartículas "virtuales" que se están materializando constantemente en parejas, alejándose y acercándose para aniquilarse de nuevo entre sí. Estas partículas se llaman virtuales porque, a diferencia de las "reales", no pueden observarse directamente en el detector de partículas. Sin embargo, puede medirse sus efectos indirectos, y su existencia ha sido confirmada por un pequeño corrimiento ("efecto Lamb") que producen en el espectro de luz de los átomos de hidrógeno excitados. En presencia de un agujero negro, un componente de un par de partículas virtuales puede caer en el interior del mismo, dejando al otro componente sin compañero con el cual aniquilarse. La partícula o antipartícula abandonada puede caer en el agujero negro después de su compañera, pero también puede escapar al infinito, donde aparecerá como radiación emitida por el agujero negro.

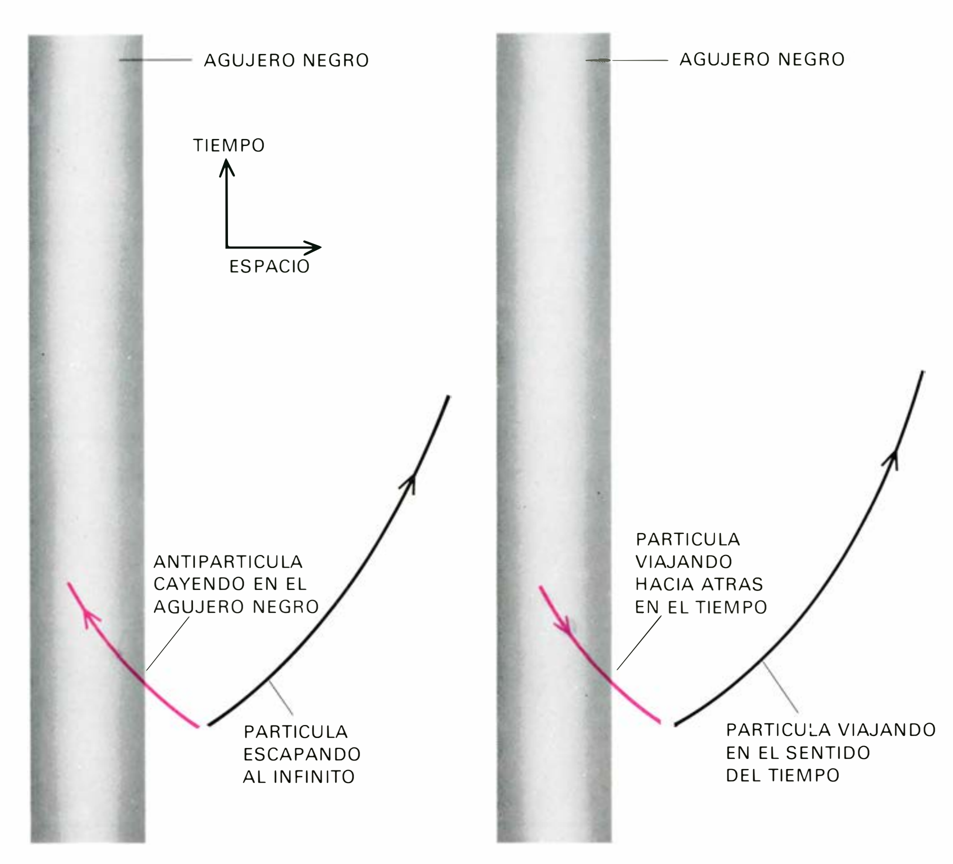
Otra manera de ver el proceso es observar al componente de la pareja de partículas que cae en el agujero negro –la antipartícula, por ejemplo– como si realmente fuera una partícula que estu-

viera viajando hacia atrás en el tiempo. Así, la antipartícula que cae en el agujero negro puede observarse como una partícula que sale del agujero y viaja hacia atrás en el tiempo. Cuando la partícula alcanza el punto en el que la pareja partícula-antipartícula se materializó originalmente, se difunde por el campo gravitatorio y viaja en el sentido del tiempo.

Por tanto, la mecánica cuántica ha permitido que una partícula escape desde el interior de un agujero negro, cosa que no era lícito en la mecánica clásica. Hay otras muchas situaciones, en física atómica y nuclear, en las que se da algún tipo de barrera que las partículas no podrían traspasar de cumplirse los principios clásicos, pero que pueden hacerlo, por efecto túnel, según los principios mecánico-cuánticos.

El espesor de la barrera que limita un agujero negro es proporcional al tamaño de éste. Ello significa que muy pocas partículas podrían escapar de un agujero negro cuyo tamaño fuera el que hipotéticamente existe en Cygnus X-1, pero sí pueden abandonar rápidamente otros agujeros negros pequeños. Cálculos detallados demuestran que las partículas emitidas tienen un espectro térmico correspondiente a una temperatura que aumenta en función directa con la reducción de la masa de un agujero negro. Para un agujero negro con la masa del sol, la temperatura es tan sólo la diezmillonésima parte de grado por encima del cero absoluto. La radiación térmica que sale de un agujero negro con esta temperatura quedaría completamente celada por la radiación de fondo general del universo. Por otra parte, un agujero negro con una masa de sólo mil millones de toneladas, esto es, un agujero negro primario de aproximadamente el tamaño de un protón, tendría una temperatura de unos 120 mil millones de grados Kelvin, que corresponden a una energía de 10 millones de electrón-voltios. A esa temperatura, un agujero negro podría crear pares electrón-positrón y partículas de masa nula, tales como fotones, neutrinos y gravitones (los presuntos portadores de energía gravitatoria). Un agujero negro primario liberaría energía a un ritmo de 6000 megawatios, equivalente al rendimiento de seis grandes centrales de energía nuclear.

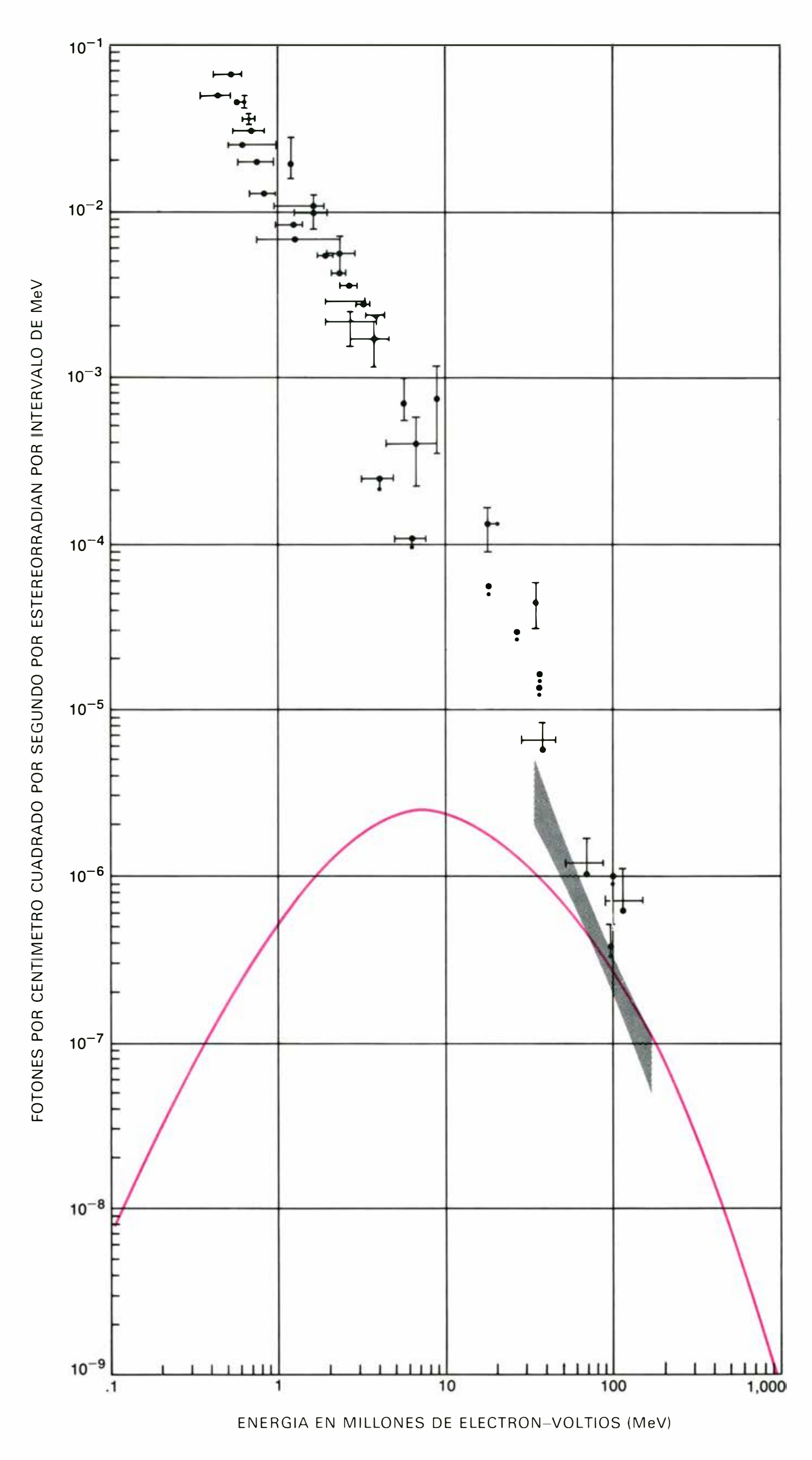
A medida que un agujero negro va emitiendo partículas, su masa y tamaño disminuyen continuamente. Esto hace que sea más fácil para un mayor número de partículas salir por efecto túnel y que la emisión continúe a un ritmo creciente, hasta que, tal vez, deje de existir. Con el tiempo, todo agujero negro del universo se evaporará de esta manera. Para agujeros negros grandes, el tiempo en que esto ocurrirá será de máxima duración: un agujero negro, cuya masa equivalga a la del sol, tardará alrededor de 10^{66} años. Así, un agujero negro primario se habría evaporado casi completamente en los 10 mil millones de años transcurridos desde la gran explosión, es decir, desde el comienzo del universo según lo conocemos. Tales agujeros negros estarían ahora emitiendo rayos gamma con una energía de unos 100 millones de electrón-voltios.



DOS INTERPRETACIONES DISTINTAS cabe dar a propósito de la emisión de partículas por un agujero negro. Una de ellas *(izquierda)* recurre a la formación de un par virtual partícula-antipartícula, un componente del cual es atrapado por el agujero negro mientras que el otro escapa del mismo. De acuerdo con la otra *(derecha)*, puede considerarse a una antipartícula que cae en un agujero negro como si fuera una partícula normal que viaja hacia atrás en el tiempo, y que va saliendo del agujero negro. Una vez fuera, se difunde por el campo gravitatorio, y se convierte en una partícula que viaja en el sentido del tiempo y escapa al infinito.

Unos cálculos hechos por Don N. Page, del Instituto de Tecnología de California y por el autor, basados en mediciones de la radiación gamma del fondo cósmico sacadas por el satélite SAS-2, demuestran que la densidad promedio de los agujeros negros primarios en el universo debe ser inferior a los 200 por año-luz cúbico. La densidad local de nuestra galaxia podría ser un millón de veces mayor que en el modelo anterior, si los agujeros negros primarios estuvieran concentrados en el "halo" de galaxias –la delgada nube de estrellas de rápido movimiento en las que está sumergida cada galaxia– en lugar de tener una distribución uniforme a través del universo. Esto implicaría que el agujero negro primario más cercano a la tierra se halle probablemenre a una distancia parecida a la que se encuentra el planeta Plutón.

La etapa final de la evaporación de un agujero negro transcurriría de una forma tan rápida que finalizaría en una enorme explosión. La potencia de esta explosión dependería de cuántas clases diferentes de partículas elementales hubiera. Si, como es ampliamente aceptado, todas las partículas constan, verosímilmente, de seis clases diferentes de quarks, la explosión final tendría una energía equivalente a unos 10 millones de bombas de hidrógeno de un megatón. Por otra parte, una teoría distinta de partículas elementales, avanzada por R. Hagedorn, de la Organización Europea para la Investi-



LOS AGUJEROS NEGROS PRIMARIOS, cuyo tamaño fuera el de una partícula elemental y un peso próximo a mil millones de toneladas, pudieron formarse, en gran número, poco después de la gran explosión, que es el origen conocido del universo. Estos agujeros negros tendrían una temperatura de unos 70 mil millones de grados Kelvin, que corresponden a una energía de 10 millones de electrón-voltios (MeV). Las partículas emitidas con esta energía producirían un espectro difuso de rayos gamma detectable por los satélites. Los datos y la región sombreada representan mediciones efectivas del espectro difuso de rayos gamma en el espacio inmediato. Las mediciones indican que la densidad media de tales agujeros negros en el universo tiene que ser inferior al millón por año-luz cúbico. La curva continua es el espectro previsto.

gación Nuclear, argumenta que hay un número infinito de partículas elementales de masas cada vez mayores. A medida que un agujero negro se va haciendo más pequeño y más caliente, iría emitiendo un número cada vez mayor de tipos de partículas diferentes y produciría una explosión unas 100.000 veces más potente que la calculada de acuerdo con la hipótesis de los quarks. De ahí que la observación de una explosión de un agujero negro proporcionaría información muy importante sobre la física de partículas elementales, información que quizá no pueda obtenerse por otros métodos.

La explosión de un agujero negro produciría una erupción masiva de rayos gamma de alta energía. Aunque podrían ser observados con detectores de rayos gamma en satélites o globos, sería difícil enviar un detector, suficientemente capaz, para así tener una probabilidad razonable de interceptar un número significativo de rayos gamma procedentes de una explosión. Una posibilidad sería emplear una plataforma espacial para construir un gran detector de rayos gamma en órbita. Otra, mucho más sencilla y barata, sería utilizar la capa superior de la atmósfera de la tierra como detector.

Un rayo gamma de alta energía sumergido en la atmósfera crearía una cascada de pares electrón-positrón, que, inicialmente, viajarían a una velocidad superior a la luz cuando atraviesa la atmósfera (la luz es frenada por las interacciones con las moléculas del aire). De este modo, los electrones y positrones formarían una especie de explosión sonora, u onda de choque, en el campo electromagnético. Tal onda de choque, llamada radiación de Cerenkov, sería detectada desde la tierra como un destello de luz visible. El experimento preliminar hecho por Neil A. Porter y Trevor C. Weekes, del University College de Dublin, indica que, si un agujero negro explota de la forma que prevé la teoría de Hagedorn, hay menos de dos explosiones de agujeros negros por año-luz cúbico por siglo en nuestra región de la galaxia. Esto implicaría que la densidad de los agujeros negros primarios no llega a los 100 millones por año-luz cúbico. Se podía incrementar grandemente la sensibilidad de tales observaciones. Aunque no dieran ninguna prueba positiva de agujeros negros primarios, resultarían muy valiosas. Estableciendo un límite superior bajo en la densidad de tales agujeros negros, las observaciones pondrían de manifiesto

que, al principio, el universo tuvo que ser muy calmado, sin turbulencias.

La gran explosión parece la explosión de un agujero negro a gran escala. De ahí que si se logra comprender como los agujeros negros crean partículas, se llegará a un entendimiento similar de como la gran explosión lo originó todo en el universo. En un agujero negro, la materia se colapsa y se pierde para siempre, pero, en su lugar, se crea nueva materia. Por ello, pudo darse una fase inicial del universo durante la cual se colapsó la materia, para ser creada de nuevo en la gran explosión.

Si la materia que se colapsa para formar un agujero negro tiene una carga eléctrica neta, el agujero negro resultante tiene la misma carga. Esto significa que el agujero negro tenderá a atraer a los componentes de los pares virtuales partícula-antipartícula que tengan carga opuesta y a repeler a los que tengan la misma carga. Por esta razón, el agujero negro emitirá preferentemente partículas con carga de su mismo signo y perderá así rápidamente su carga. Análogamente, si la materia que se colapsa tiene un momento angular neto, el agujero negro resultante girará y emitirá preferentemente partículas que lleven su momento angular. La razón de que un agujero negro "recuerde" la carga eléctrica, el momento angular y la masa de la materia que colapsó y "olvide" todo lo demás radica en que estas tres cantidades se hallan asociadas a campos de largo alcance: en el caso de la carga, el campo electromagnético, y, en el caso del momento angular y de la masa, el campo gravitatorio.

Los experimentos llevados a cabo por Robert H. Dicke, de la Universidad de Princeton, y Vladimir Braginsky, de la Universidad estatal de Moscú, han concluido que no hay campos de largo alcance asociados con la propiedad cuántica denominada número bariónico (los bariones son la clase de partículas que abarca el protón y el neutrón). De aquí que un agujero negro formado por el colapso de un conjunto de bariones prescindiría de su número bariónico y radiaría igual cantidad de bariones que de antibariones. Por ello, cuando el agujero negro desapareciera, se habría violado una de las leyes más apreciadas de la física de partículas, la ley de la conservación del número bariónico.

Aunque la hipótesis de Bekenstein, sobre la entropía finita de los agujeros negros, requiere, para no ser contradictoria, que los agujeros negros radien térmicamente, en principio parece un completo milagro que los minuciosos cálculos mecánico-cuánticos de la creación de partículas den la emisión con un espectro térmico. La explicación es que las partículas emitidas salen del agujero negro desde una región de la cual un observador externo sólo conoce su masa, momento angular y carga eléctrica. Esto significa que todas las combinaciones o configuraciones de partículas emitidas, que tienen la misma energía, momento angular y carga eléctrica, son igualmente probables. Incluso es posible que el agujero negro pueda emitir una serie de televisores o las novelas de Proust en diez volúmenes encuadernados en piel, pero el número de configuraciones de partículas que corresponden a estas exóticas posibilidades es despreciablemente pequeño. El número mayor, y con mucho, de configuraciones corresponde a la emisión con un espectro que es aproximadamente térmico.

La emisión desde un agujero negro tiene un grado adicional de incertidumbre, o impredictibilidad, muy por encima del asociado normalmente con la mecánica cuántica. En la mecánica clásica, se puede predecir la medida tanto de la posición como de la velocidad de una partícula. En la mecánica cuántica, el principio de incertidumbre dice que se puede predecir únicamente una de estas dos cantidades; el observador puede predecir el resultado de medir la posición o la velocidad, pero no ambas. Alternativamente, puede predecir el resultado de medir una combinación de posición y velocidad. Por consiguiente, la capacidad que tiene el observador de hacer predicciones definidas queda reducida a la mitad. Con los agujeros negros la situación es aún peor. Como las partículas emitidas por un agujero negro proceden de una región de la que el observador tiene un conocimiento muy limitado, no puede predecir de una manera exacta la posición y la velocidad de una partícula o una combinación de las dos: todo lo más que puede predecir es la probabilidad de que se emitan ciertas partículas. Al parecer, pues, Einstein estaba doblemente equivocado cuando afirmó que "Dios no juega a los dados". Los estudios sobre la emisión de partículas desde agujeros negros permiten sospechar que Dios no solamente juega a los dados, sino que, a veces, los echa donde nadie puede verlos.

Ciencia y sociedad

El control del material hereditario

Dos poderosas técnicas de la genética molecular contemporánea –la técnica del ADN recombinante y la síntesis del ADN *in vitro*– se han conjugado hace poco, por vez primera, en un empeño por entender las secuencias de ADN denominadas genes reguladores, que controlan la expresión de los genes estructurales adyacentes dentro de la unidad genética funcional conocida como operón. En sendos artículos de *Nature,* dos grupos independientes, el uno encabezado por Ray J. Wu, de Cornell University, y Saran A. Narang, del National Research Council de Canadá, y el otro por Herbet W. Boyer, de la Universidad de California (San Francisco), Richard E. Dickerson, del Instituto de Tecnología de California, y Arthur D. Riggs, del City of Hope Medical Center de Duarte, California, describen el éxito logrado en la producción de grandes cantidades de un gene regulador sintético, para análisis estructural, mediante la generación de muchas copias de dicho gene dentro de las células vivas de la enterobacteria *Escherichia coli.*

El gene investigado por ambos grupos fue el operador *lac,* situado en el cromosoma de *E. coli* junto a tres genes estructurales que codifican la síntesis de enzimas participantes en el metabolismo del azúcar lactosa. Al igual que ocurre con otros genes reguladores, el operador *lac* no se transcribe en el ARN mensajero; en cambio, controla la transcripción de los genes estructurales próximos sirviendo como *locus* de reconocimiento de una proteína denominada represor *lac.* En ausencia de lactosa, el represor (del que hay unas 10 copias por célula) se une a la secuencia operador, bloqueando físicamente la transcripción de los genes estructurales *lac* mediante el enzima ARN polimerasa. Sin embargo, en presencia de lactosa, una forma modificada del azúcar se liga al represor, cambiando su forma de manera que no pueda reconocer al ADN operador. Cuando todas las moléculas represoras han sido inactivadas de modo análogo, el ARN polimerasa puede unirse al ADN en un *locus* adyacente al operador denominado promotor, y los tres genes estructurales *lac* se transcriben de forma secuencial.

El mecanismo mediante el cual el represor *lac* reconoce y se asocia específicamente con el gene operador *lac,* de entre miles de otras secuencias de nucleótidos de la molécula de ADN, presenta el mayor interés, pero su elucidación exigirá la cristalización y el análisis de complejos represor-operador, tarea que supone grandes cantidades de ADN operador purificado (ya se dispone de represor *lac* puro). Con este fin, se ha determinado la secuencia de 21 pares de nucleótidos del operador *lac* y se le ha sintetizado, pero en el laboratorio sólo se han obtenido escasas cantidades del gene puro. A objeto de multiplicar el rendimiento del gene operador sintético, los grupos de California y de Cornell idearon una nueva aplicación de la tecnología de recombinación del ADN. Utilizando la notable especificidad de los enzimas restrictivos, que reconocen una secuencia palindrómica singular de nucleótidos en la espiral de ADN y hacen cortes discontinuos en ambas cadenas de la espiral siempre que aparece esta secuencia, sintetizaron la conocida secuencia de reconocimiento para el enzima de restricción designado *EcoRI* y agregaron una copia a cada extremo del gene operador sintético. El mismo enzima de restricción se usó entonces para hacer voladizos de un segmento en un plásmido, que es un trozo circular de ADN no cromosómico que puede trasladarse fácilmente de una bacteria a otra.

Como los voladizos sintetizados de un segmento añadidos al ADN operador *lac* coincidían con los creados por el enzima de restricción en el ADN de plásmido, estas regiones se aparearon espontáneamente al incubar conjuntamente ambas moléculas. Se agregó luego el enzima polinucleótido ligasa para unir covalentemente las estructuras de azúcar-fosfato de los dos ADN, de suerte que el gene operador sintetizado quedaba empalmado en el plásmido. Cuando los híbridos plásmido-operador se introdujeron en células de *E. coli,* procedieron a reproducirse numerosas veces.

Las colonias bacterianas que contenían el ADN híbrido (alrededor del 3 por mil del total) se identificaron según dos criterios: primero, eran resistentes al antibiótico tetraciclina, en virtud de un gene aportado por el plásmido, y, segundo, las 30 y tantas copias del operador *lac* producidas por la multiplicación del plásmido se acoplaron rápidamente las 10 moléculas de represor *lac* de cada célula, impidiendo de ese modo que el represor se ligase al único gene operador *lac* del cromosoma de *E. coli.* El resultado fue que los genes estructurales *lac* del huésped fueron transcritos y las bacterias afectadas empezaron a fabricar los tres enzimas necesarios para el metabolismo de la lactosa, aunque no hubiese presencia de ésta en el medio. Uno de estos enzimas, la beta-galactosidasa, fue detectado por el grupo de California al alimentar las bacterias con un análogo de la lactosa que contenía una molécula del pigmento índigo. El azúcar sintético carecía de todo efecto inductor sobre el represor, pero sirvió como sustrato para la beta-galactosidasa. Al ser atacado por el enzima, liberó el pigmento, tiñendo de azul esas colonias.

De las colonias azules se extrajo entonces los híbridos operador-plásmido multicopiador y se les digirió con enzima de restricción, y así quedaron divididos los "vinculadores" de la secuencia de restricción de cada lado del operador *lac.* Luego se liberó el ADN operador del ADN plásmido, con un rendimiento de muchas veces la cantidad original de gene sintético. Aún así, el mayor rendimiento producido por la multicopia ("cloning") de secuencias singulares de operador seguía siendo demasiado escaso para experimentos fisicoquímicos, y ambos grupos de científicos están intentando producir plásmidos que contengan copias múltiples de operador *lac* sintético. Wu y sus colaboradores informan que han usado hace poco la polinucleótidoligasa para ligar hasta 18 genes operadores en tándem, con vinculadores sintéticos de secuencia de restricción entre ellos.

Las posibilidades de esta técnica se extienden obviamente más allá de su

empleo actual. En principio, los vinculadores de la secuencia de restricción podrían añadirse a cualquier trozo de ADN sintético y la secuencia deseada podría multicopiarse entonces en plásmidos en grandes cantidades. Más aún, Wu y sus colegas proponen que podría introducirse un operador *lac* sintético frente a un determinado gene o grupo de genes, permitiendo al experimentador activar o inactivar estos genes con el represor *lac*. Así, la síntesis química y la técnica de multicopia conjugadas permiten al genetista molecular fabricar cualquier secuencia de ADN que desee, producirla en cantidades ilimitadas y finalmente insertarla a cualquier otro trozo de ADN que elija. Por primera vez el hombre ha desarrollado la capacidad de control casi absoluto sobre el material de los genes.

Ambigüedades de Marte

La pausa en la riada de datos provenientes de las cápsulas *Viking* dejadas en Marte, inevitable por ponerse dicho planeta tras el sol a fines de noviembre, dio oportunidad a los investigadores de reflexionar sobre los resultados de las observaciones tomadas desde los dos vehículos espaciales, el primero de los cuales fue depositado, ileso, en la superficie marciana el 20 de julio pasado y el segundo el 3 de septiembre. La cápsula del *Viking I* se detuvo en la depresión de Chryse, a unos 20 grados al norte del ecuador, región baja en la que parecen haberse avenado varios canales. La cápsula del *Viking II* se detuvo en la llanura de Utopía, a casi 180 grados de la primera cápsula y a unos 28 grados más al norte. La región de Utopía, que constituye una parte de los vastos llanos que ocupan gran parte del hemisferio norte de Marte, es un terreno volcánico. Los datos procedentes de los *Viking*, que se remontan hasta el momento de la conjunción del planeta con el sol, están resumidos en una serie de estudios aparecidos en *Science*.

Pese a su distinta peculiaridad geológica, los dos lugares de "amartizaje" se ven notablemente parecidos en las fotos tomadas en la superficie. Ambos lugares, sembrados de rocas, son de color rojo anaranjado; el cielo se ve rosado. El brillante color de óxido del paisaje se debe a un fino revestimiento de óxido de hierro (quizá goethita o hematita) sobre las rocas y partículas superficiales. El análisis de las partículas superficiales sugiere que sus principales componentes son un 45 por ciento de óxido de silicio (SiO_3), un 18 por ciento de óxido de hierro (Fe_2O_3), junto con un 5 a un 8 por ciento, cada uno, de óxidos de aluminio, calcio, magnesio y azufre. El contenido de azufre es de 10 a 100 veces mayor que el de los suelos terrestres o lunares típicos, y más o menos igual al de los meteoritos condríticos, que, según se cree, representan los materiales primitivos de los cuales se formaron los planetas interiores. En conjunto, la composición del suelo marciano es diferente del de cualquier tipo simple o abundante de roca o suelo hallado en la Tierra o en la Luna. Parece por tanto que Marte no ha llegado a diferenciarse químicamente tanto como la Tierra o la Luna.

Tal como se esperaba, las observaciones de las cápsulas muestran que la atmósfera marciana contiene 95 por ciento de anhídrido carbónico y menos de 0,5 por ciento de oxígeno. Pero el hallazgo de 2 a 3 por ciento de nitrógeno echa por tierra la hipótesis de que este gas pudiera faltar enteramente. La composición total de la atmósfera sugiere que los gases ocluidos en el interior de Marte quizá hayan sido liberados más lentamente que los gases terrestres. El clima en ambos lugares de amartizaje se describe como "repetitivo y suave". La temperatura diaria fluctúa entre -85 y -30 grados Celsius. Se ha registrado ráfagas de viento hasta de 68 kilómetros por hora, demasiado lentas como para levantar mucho polvo.

Los tres experimentos ideados para detectar vida en Marte han dado resultados sugestivos, pero no concluyentes. El más general de los tres, el experimento de intercambio gaseoso, mide los cambios de composición a largo plazo en el gas que sobrenada una muestra de suelo después de haber añadido un medio nutritivo acuoso. Las muestras han liberado cantidades considerables de anhídrido carbónico y oxígeno, y cantidades menores de nitrógeno. Si bien es posible que el anhídrido carbónico sea producido por un proceso biológico de oxidación, parece mucho más probable la producción inorgánica del gas.

El experimento de asimilación de carbono, o de liberación pirolítica, mide la capacidad del suelo marciano para convertir dióxido o monóxido de carbono, marcado con el isótopo radiactivo carbono 14, en materia orgánica. La muestra del suelo se coloca en una cámara de ensayo que contiene una atmósfera de los dos gases marcados y se expone a luz solar marciana simulada, es decir, luz rica en radiación ultravioleta. A las 120 horas, se extrae la atmósfera de la cámara y se calienta el suelo a 625 grados C para volatizar cualquier compuesto orgánico que se haya formado. Los productos orgánicos, junto con el monóxido y dióxido de carbono sin reaccionar, se trasladan a una columna que los separa en dos fracciones: fracción 1, que contiene monóxido de carbono, dióxido de carbono y metano (CH_4), y fracción 2, que contiene todo fragmento molecular orgánico de mayor tamaño que el metano. Así pues, la fracción 2 representaría los compuestos carbónicos sintetizados en suelo marciano. Tres muestras del primer lugar de amartizaje, en Chryse, dieron recuentos significativamente altos para la fracción 2. Una cuarta muestra, cuya temperatura subió 10 grados más de lo previsto durante la incubación (26° C en vez de 15° a 17° C) mostró una fracción débilmente positiva. Una muestra de control del suelo de Chryse, esterilizada por calor antes de iniciar la incubación, no mostró ningún recuento significativo en la fracción 2.

En el emplazamiento de Utopía se ensayaron tres muestras. Una, incubada a oscuras, dio un resultado débilmente positivo. Otra segunda, incubada con luz pero añadiendo vapor de agua a la atmósfera experimental, dio un resultado claramente negativo. Una tercera muestra, tomada de debajo de una roca e incubada también a oscuras, dio también resultados negativos. Los investigadores llegaron a la conclusión de que "en el material de la superficie de Marte se produce una síntesis orgánica a partir del CO o del CO_2 atmosféricos. La síntesis es débil comparada con la que se observa en suelos terrestres biológicamente activos y, a diferencia de éstos, es inhibida por pequeñas cantidades de agua. Se parece a las reacciones biológicas en que es termolábil, aunque hay motivos para dudar de que la síntesis quedase totalmente anulada en la muestra (esterilizada) de Chryse".

En el tercer ensayo en busca de la vida marciana, que es el experimento de liberación de nutrientes marcados, se intenta detectar la presencia de metabolismo, con o sin crecimiento, vigilando la evolución de gas que contiene carbono 14, a partir de una muestra de suelo a la cual se ha agregado un nutriente compuesto de siete substratos orgánicos (formato, glucolato, glicina, D-alanina, D-lactato y L-lactato) marcados con carbono 14. La oxidación del carbono en cualquiera de las 17 posiciones moleculares marcadas daría por consecuencia un gas que generase unos 15.000 recuentos de carbono 14 por minuto. Hasta hoy se ha llevado a cabo dos ensayos con

suelo activo y uno con suelo esterilizado por calor en cada uno de los emplazamientos de amartizaje. Las cuatro pruebas con suelo activo liberaron gas cada una, a razón de 10.000 a 14.000 recuentos por minuto. Los dos ensayos con suelo esterilizado (calentado a 160º C en un ensayo y a 50ºC en otro) mostraron poca o ninguna actividad por encima de los niveles de fondo respecto del gas liberado. Los investigadores advierten que los resultados podrían provenir de substancias inorgánicas del suelo marciano, que hubiesen sido activadas por exposición a la radiación ultravioleta. "Pero, hasta la fecha, ningún experimento químico ha reproducido cuantitativamente los datos (de liberación de carbono marcado) de Marte. Por tanto, sigue existiendo la posibilidad de que se haya registrado actividad biológica en Marte".

Un cuarto experimento, no proyectado explícitamente para indagar la existencia de vida, ha sido tal vez el que más ha alertado a los investigadores a la hora de interpretar los tres experimentos de biología. Se trata del análisis de muestras de suelo mediante una combinación de cromatografía gaseosa y espectrometría de masas. El ensayo GCMS, según se le denomina, puede detectar la presencia de compuestos orgánicos a razón de unas pocas partes por cada mil millones. Lo sorprendente es que en las muestras marcianas no se ha hallado compuesto orgánico de ninguna especie. Sólo se ha detectado agua, anhídrido carbónico y trazas de contaminantes terrestres conocidos. Se podría contar con que los organismos marcianos, si existiesen, dejaran tras sí un desecho mensurable de materia orgánica. Por ejemplo, en una muestra de suelo de la Antártida, que según se cree contiene menos de 10.000 microorganismos por gramo, se halló más de 20 compuestos orgánicos en concentraciones fluctuantes entre una parte por millón y una por 100 millones.

Para esta ausencia de compuestos orgánicos en Marte se ha propuesto varias explicaciones. Si el suelo contuviera no más de, digamos, 1000 microorganismos por gramo, sus constituyentes no serían detectados en el experimento GCMS. Además, los organismos muertos podrían no acumularse hasta niveles detectables si fueran altamente biodegradables en el entorno marciano de fuerte radiación ultravioleta, conjugada con catalizadores de óxido metálico existentes en el suelo. Puede que una serie de nuevos experimentos actualmente en curso en Chryse y en Utopía resuelvan estas ambigüedades.

El lado de unión de los anticuerpos

Cuando un antígeno, o una sustancia extraña, se combina con una molécula del anticuerpo, lo hace en un sitio donde encaja exactamente, como encaja la llave en su cerradura. La naturaleza de este lado de unión se ha descrito con detalle

J. Donald Capra y Allen B. Edmundson

Cuando las bacterias o los virus atacan al cuerpo, éste tiene dos maneras principales de defenderse. En primer lugar, los glóbulos blancos de la sangre pueden trasladarse hacia el punto de ataque para englobar y destruir a los agentes extraños. En segundo lugar, unas células, conocidas como células plasmáticas, pueden fabricar anticuerpos, que son unas proteínas altamente específicas y a las cuales la corriente sanguínea transporta para inutilizar a los invasores. La parte del organismo atacante que es responsable de estimular la producción de anticuerpos se llama antígeno; suele tratarse de un pequeño fragmento de la superficie de una gran molécula de proteína o de hidrato de carbono asociada con la estructura externa del microorganismo. Una bacteria, e incluso una gran molécula proteica, tal como una toxina bacteriana, puede disponer de varios sitios antigénicos de éstos en su superficie, y, en consecuencia, puede inducir la producción de varios anticuerpos específicos distintos. Una vez liberados en el interior del torrente sanguíneo, los anticuerpos se combinarán con sus correspondientes antígenos, como si fueran piezas complementarias de un rompecabezas, modificando de este modo la estructura externa del microorganismo invasor para que un leucocito pueda ingerirlo más fácilmente. Por su parte, los anticuerpos solos pueden neutralizar microorganismos o toxinas aglutinándose con ellos en grandes complejos inactivos.

LOS ANTICUERPOS NEUTRALIZAN A LOS VIRUS INVASORES ligándose específicamente a pequeños lados antigénicos de la superficie externa del virus. Cada molécula de anticuerpo tiene dos lados de unión idénticos, de modo que puede unirse a regiones adyacentes de la misma partícula vírica o entrelazar dos partículas víricas formándose un complejo inactivo. En esta micrografía electrónica, las moléculas del anticuerpo forman unos halos tenues en torno a las partículas esféricas de virus del polioma, de 400 angstroms de diámetro. Uniéndose a varios antígenos en un solo virus, los anticuerpos alteran la estructura de éste para que pueda ser más fácilmente englobado por los leucocitos sanguíneos. Micrografía de Almeida (Ontario Cancer Institute) y Cinader-Howatson (Universidad de Toronto).

Aunque los anticuerpos son absolutamente esenciales para la supervivencia de los individuos y de la especie, resultan también peligrosos en potencia. Si no son altamente específicos para antígenos extraños, podrían atacar a las propias moléculas del cuerpo, muchas de las cuales se parecen estrechamente a las moléculas de los microorganismos invasores. Algo así sucedería realmente en los desórdenes conocidos como enfermedades de autoinmunidad; en éstas, el cuerpo pierde aparentemente la capacidad de distinguir entre lo "propio" y lo "extraño" y produce anticuerpos que actúan contra sus propias proteínas o sus ácidos nucleicos, causando a menudo graves daños en los tejidos que, algunas veces, resultan fatales.

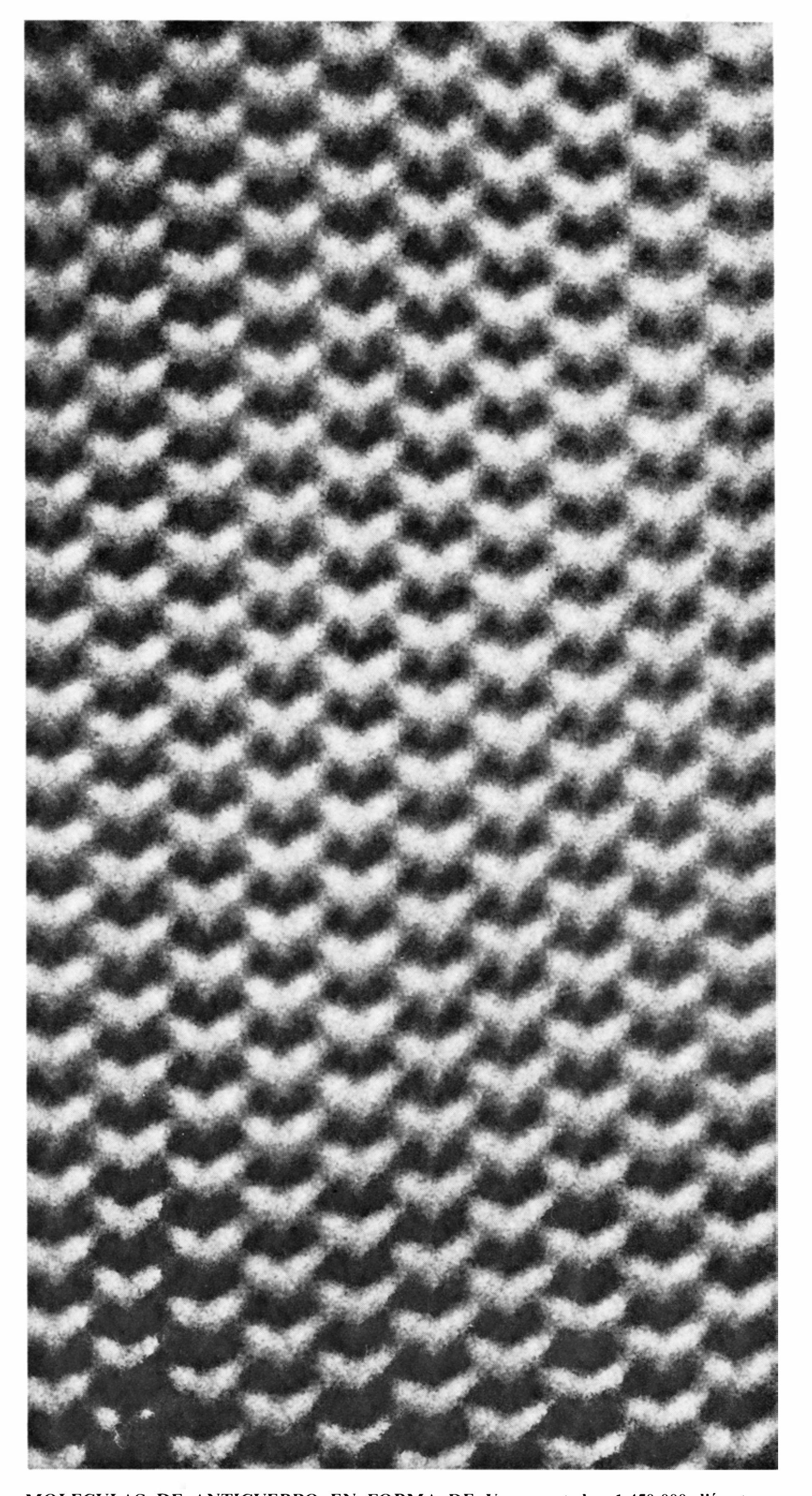
El éxito evolutivo de los anticuerpos ha requerido, por lo tanto, que fueran

altamente sensibles al menor cambio estructural en los antígenos que reconocen. Esto ya se indicó hace unos 50 años, cuando Karl Landsteiner, del Rockefeller Institute for Medical Research, descubrió que los anticuerpos podían distinguir entre isómeros ópticos, es decir, entre moléculas con idéntica composición química que solamente difieren en la distinta colocación, en el espacio tridimensional, de algunos átomos. Más recientemente, la aplicación de los conceptos y los métodos de la biología molecular a la estructura tridimensional del lado de unión del anticuerpo ha posibilitado nuevos y valiosos conocimientos acerca de la impresionante especificidad con que los anticuerpos pueden actuar.

Estructura de la molécula del anticuerpo

Los anticuerpos pertenecen a la clase de proteínas llamadas inmunoglobulinas. La inmunoglobulina gamma (designada como IgG) es el anticuerpo predominante en la sangre de la mayoría de los vertebrados superiores, y ha sido, de los cinco tipos de inmunoglobulinas, la que se ha estudiado con mayor intensidad. Tiene un peso molecular aproximado de 160.000, lo cual indica que su molécula consta de unos 23.000 átomos. A principios de la década de los 60, dos grupos de investigadores, dirigido uno por Rodney R. Porter, en Inglaterra, y otro por Gerald M. Edelman, en la Rockefeller University, demostraron que la IgG está formada por cuatro cadenas polipeptídicas, cada una de las cuales consta de unidades de aminoácidos enlazados por enlaces peptídicos. Las cuatro cadenas están apareadas, de modo que la molécula consta de dos mitades idénticas, cada una con una cadena "pesada" o larga y una cadena "ligera" o corta. Las dos cadenas de cada par están unidas entre sí por enlaces covalentes entre los átomos de azufre de los aminoácidos cistina. Si estos enlaces disulfuro se rompen, las cadenas pesada y ligera permanecerán unidas entre sí por interacciones no covalentes. Cuando se introducen en una solución de ácido o urea, se disocian, sin embargo, en cadenas pesadas y ligeras, las cuales pueden separarse por su diferente tamaño.

Experimentos llevados a cabo por el grupo de Porter permitieron una descripción de la molécula del anticuerpo desde el aspecto funcional. Cuando las moléculas de IgG se trataban con el enzima proteolítico papaína, la molécula se dividía en tres partes de tamaño aproximadamente igual. Dos de ellas eran idénticas y se designaron como Fab (de



MOLECULAS DE ANTICUERPO EN FORMA DE Y, aumentadas 1.450.000 diámetros. Esta micrografía electrónica corresponde a un cristal de inmunoglobulina humana, obtenida de un sujeto que padecía mieloma múltiple. La interferencia que resultó de la difracción del haz de electrones por el cristal produjo un negativo fotográfico de baja resolución; al sacar un positivo, se mejoró la resolución por medio de una técnica promediadora. L. W. Labaw y D. R. Davies, del National Institute of Health, realizaron una breve exposición, movieron el papel una unidad y lo volvieron a exponer. Repitiéndolo varias veces, lograron esta imagen.

"fragment, antigen-binding", fragmento de unión con el antígeno). Cada fragmento Fab de una molécula de IgG posee un lado de unión. La molécula intacta de anticuerpo tiene dos de dichos lados, por lo cual puede ligar materiales antigénicos en complejos inactivos. El tercer fragmento resultante de la digestión por la papaína se llama Fc, puesto que cristaliza rápidamente. No se une al antígeno, pero posee otras importantes actividades biológicas.

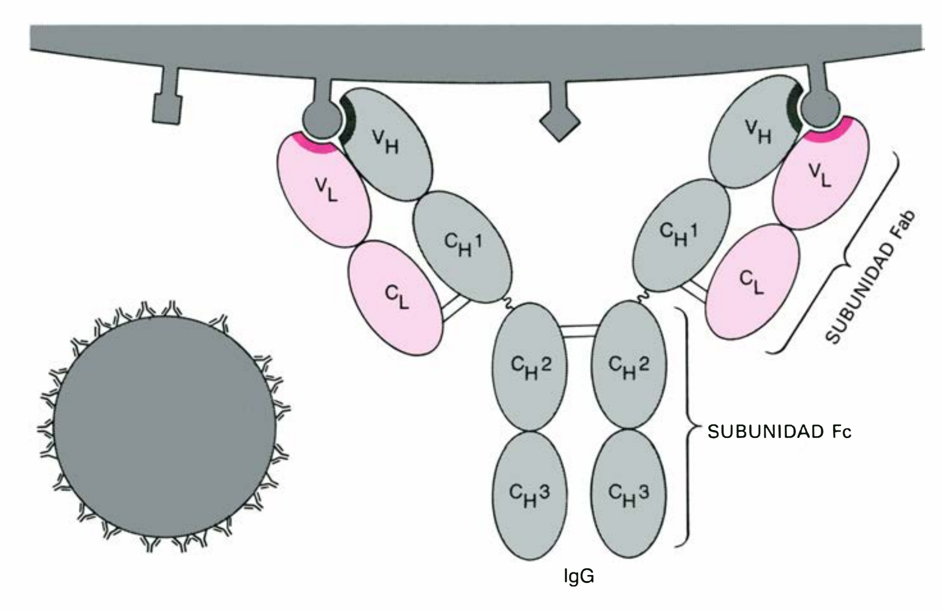
Los resultados obtenidos por muchos investigadores en los años 1960 llevaron al modelo de IgG aceptado actualmente, en el cual las cadenas pesada y ligera se disponen en forma de una *Y*. De acuerdo con este modelo, el fragmento Fc es el pie de la *Y*, y consta de la mitad inferior de las dos cadenas polipeptídicas pesadas, que están enlazadas por uno o más puentes disulfuro. Los dos fragmentos Fab son los brazos de la *Y*; cada uno consta de una cadena ligera entera y del resto de la cadena pesada, con un lado de unión para el antígeno de idéntica especificidad en el extremo. La prueba de las micrografías electrónicas apoyó este esquema e indicó también que la unidad Fc y las dos unidades Fab de la molécula intacta del anticuerpo se juntaban por medio de una bisagra que permitía variar el ángulo entre las unidades Fab. Esta geometría flexible sirve, a buen seguro, para entrelazar partículas extrañas con más eficacia que no una estructura rígida, puesto que la distancia entre los lados de unión en el anticuerpo puede adaptarse para cubrir toda una gama de distancias entre los sitios antigénicos (tales como las subunidades de proteína repetidas de un virus).

Secuencias de aminoácidos

Tras estudios con muchas proteínas, se ha podido llegar a la siguiente generalización: la secuencia de unidades de aminoácidos en la cadena polipeptídica es la que dicta la forma como ésta se pliega en una estructura tridimensional y por tanto, en el caso de un anticuerpo, la que dicta la especificidad del lado de unión. A partir de su secuencia de aminoácidos, es mucho lo que puede deducirse acerca de la estructura de la molécula del anticuerpo. En el caso de la IgG, no obstante, hay que superar una gran dificultad técnica. Las moléculas del anticuerpo, incluidas las que tienen la misma especificidad, tienden a ser ligeramente diferentes, en razón de las variaciones que se registran en su secuencia de aminoácidos. Sin embargo, estas diferencias no son lo suficientemente grandes como para permitir separar, con los métodos disponibles, moléculas muy similares de anticuerpos. Cuando esto ocurre, no puede obtenerse ninguna muestra pura que sirva para determinar la secuencia de aminoácidos ni tampoco un cristal que permita el análisis de la estructura tridimensional mediante difracción de rayos X.

El mieloma múltiple de condición cancerosa aportó la solución al problema del abastecimiento adecuado de anticuerpo puro. En esta enfermedad, que ataca al hombre y puede inducirse en ratones, una célula productora de anticuerpos se propaga ampliamente y da origen a una proliferación de células hijas. Una célula determinada, productora de anticuerpos, fabrica anticuerpos de un solo tipo, de suerte que las células del mieloma, que son un clon celular descendiente de una sola célula original, segregarán todas una molécula idéntica de inmunoglobulina. En unos casos, se acumulan en el suero moléculas enteras de IgG; en otros, se producen más cadenas ligeras que pesadas, y el exceso de cadenas ligeras se elimina por la orina, ya individualmente, ya en dímeros o pares de cadenas. Estas proteínas de la orina fueron observadas por vez primera por Henry Bence-Jones en el Guy's Hospital de Londres, en 1847; desde entonces, se conocen como proteínas de Bence-Jones. Puesto que las proteínas de Bence-Jones pueden obtenerse fácilmente de muestras de orina de pacientes de mieloma, sin necesidad de extraer sangre, fueron éstas las primeras inmunoglobulinas analizadas para la secuencia de aminoácidos.

Las primeras secuencias completas de aminoácidos de cadenas ligeras de Bence-Jones, que constan de 214 aminoácidos, quedaron determinadas en 1965 por Norbert Hilschmann y Lyman C. Craig, de la Rockefeller University, y por el grupo de Frank W. Putnam, de la Universidad de Florida. Descubrieron que las proteínas de Bence-Jones procedentes de distintos pacientes de mieloma tenían secuencias de aminoácidos distintas, y que las diferencias se limitaban principalmente a la primera mitad de la cadena polipeptídica. La cadena ligera podía dividirse, por consiguiente, en una región variable (aminoácidos del 1 al 108) y una región constante (aminoácidos del 109 al 214). Pronto, estas observaciones se extendieron a las cadenas pesadas, que constan de 446 aminoácidos. La comparación de las secuencias de las cadenas pesadas de inmunoglobulinas, procedentes de distintos mielomas, reveló que todas las diferencias en la secuencia se localizaban en 108 y hasta



ESTRUCTURA DE LA INMUNOGLOBULINA G (IgG), el más común de los cinco tipos de anticuerpos hallados en la sangre de los vertebrados superiores. IgG, que es una gran molécula proteica, consta de dos pares de cadenas polipeptídicas, ligeras y pesadas, unidas por enlaces disulfuro en una estructura en forma de *Y*. Cada cadena se repliega en una serie de zonas globulares; la cadena pesada *(gris)* posee cuatro zonas de éstas y, la cadena ligera *(color)*, dos. La digestión de la molécula de IgG por el enzima proteolítico papaína dio origen a tres fragmentos: un fragmento Fc (el pie de la molécula en forma de *Y*) y dos fragmentos idénticos Fab (los brazos de la *Y*), cada uno de los cuales contiene un lado de unión de idéntica especificidad. Las dos zonas situadas en la punta de cada fragmento Fab (V_H y V_L) varían en la secuencia de aminoácidos según la especificidad del anticuerpo, por lo cual se denominan zonas variables; la mayoría de las zonas restantes de cada cadena no varían y, por ello, se llaman zonas constantes. Las cadenas laterales de aminoácidos de las zonas variables contribuyen a la formación del lado de unión que es, geométrica y químicamente, complementario de un tipo de antígeno, creando un íntimo acoplamiento cavidad-antígeno.

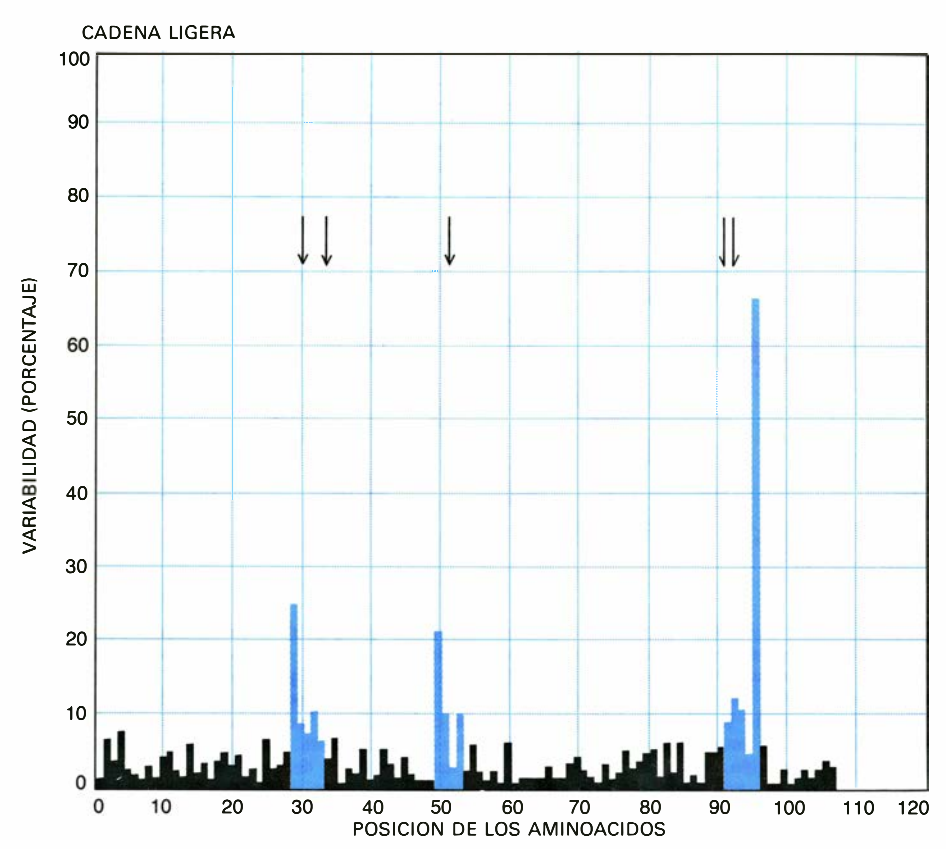
125 aminoácidos de un mismo extremo. Así, el extremo variable de la cadena pesada era aproximadamente de la misma longitud que el de la cadena ligera, mientras que el extremo constante de la cadena pesada era unas tres veces más largo.

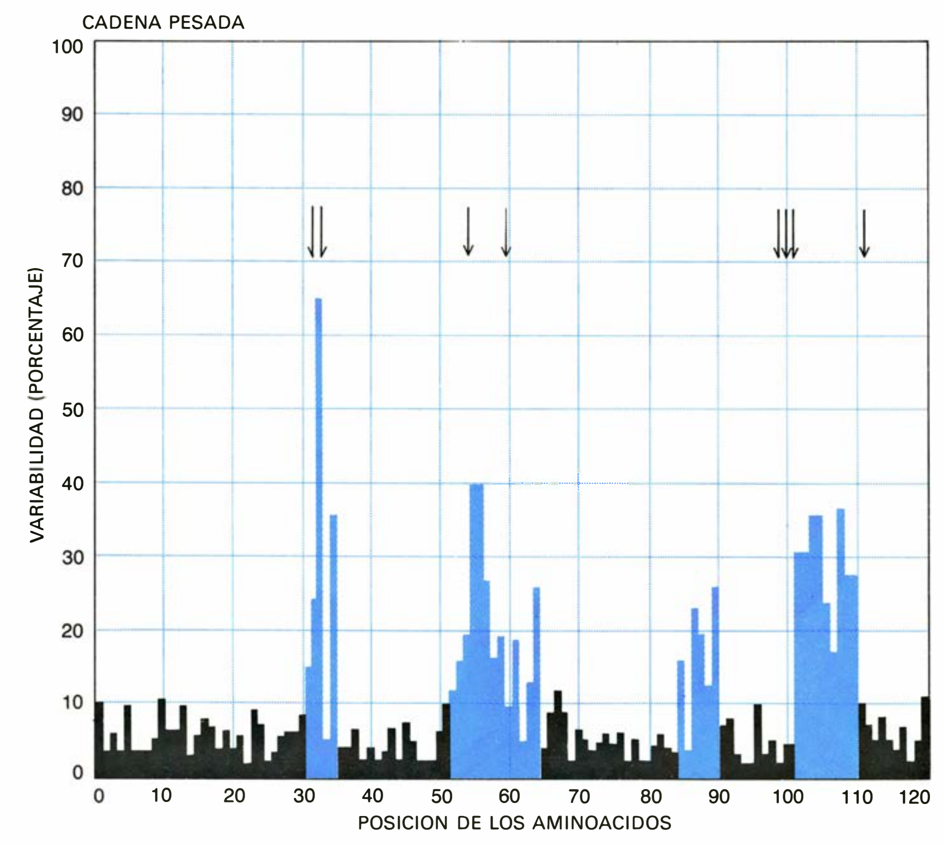
El descubrimiento de que la variación registrada en la secuencia de aminoácidos entre las moléculas de IgG se circunscribiera al mismo número aproximado de aminoácidos en el mismo extremo de las cadenas pesadas y ligeras encarrilaba decididamente la cuestión: en estas regiones era donde se unían los antígenos. Se argumentaba del siguiente modo: ya que ésta era la región en donde las distintas moléculas de IgG diferían unas de otras, tendría que ser también la región en la cual se determinaba la especificidad del anticuerpo.

En 1968, el grupo de Edelman consiguió descifrar venturosamente la secuencia completa de aminoácidos de una molécula de IgG de un mieloma. Esta secuencia mostró semejanzas sorprendentes entre segmentos de cadena polipeptídica a lo largo de partes diferentes de la molécula. La secuencia de la región variable de la cadena ligera se parecía a la secuencia de la región variable de la cadena pesada. Por otra parte, la región constante de la cadena pesada constaba de tripletes iguales, que eran de secuencia similar, y parecida a la región constante de la cadena ligera. Estas homologías en la secuencia de aminoácidos sirvió de base a la hipótesis de que la molécula de IgG está formada de territorios compactos, o subunidades, que se parecen mucho entre sí en una estructura tridimensional; hipótesis que más tarde quedaría demostrada.

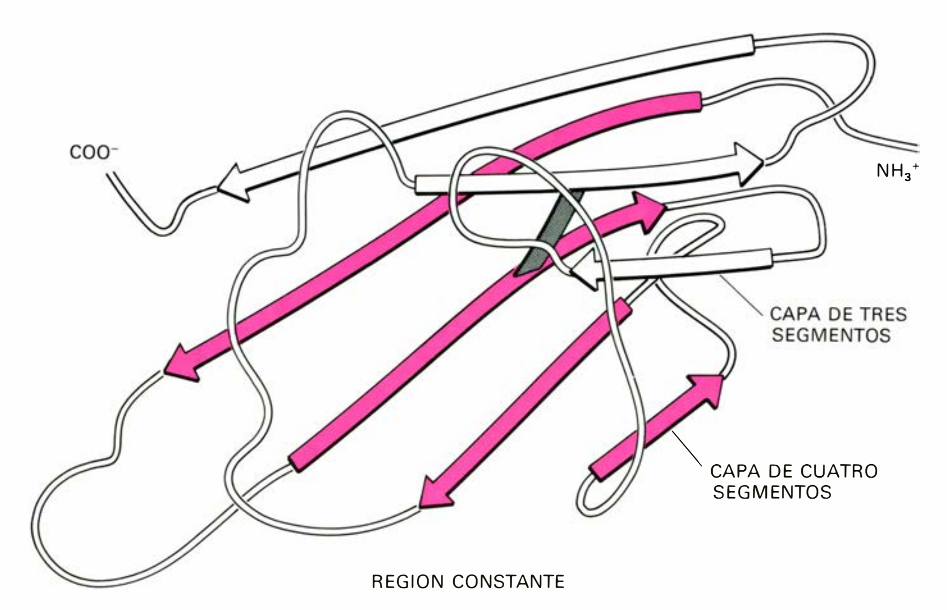
Secuencias hipervariables

En los primeros años de la década de los 70 se realizó otro avance importante. Cuando T. T. Wu y Elvin A. Kabat, de la Columbia University College of Physicians and Surgeons, analizaron las secuencias de aminoácidos de las regiones variables en cadenas ligeras de humanos y animales, observaron que tres regiones particulares de las secuencias de aminoácidos variaban más que el resto de la región variable de la cadena ligera. Uno de nosotros (Capra, trabajando a la sazón en el laboratorio de J. Michael Kehoe, en la Mount Sinai School of Medicine, de Nueva York) descubrió más tarde cuatro de tales regiones "hipervariables" en la cadena pesada. Las secuencias hipervariables abarcaban unos 25

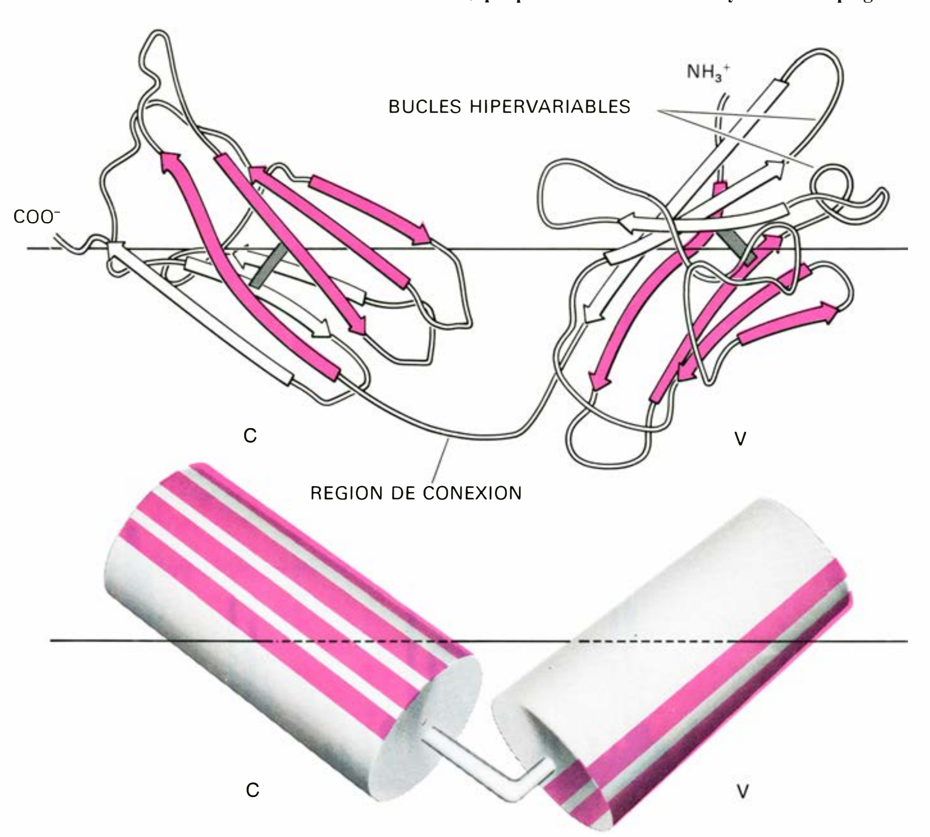




LOS AMINOACIDOS HIPERVARIABLES *(barras en color)*, **en las zonas variables de la cadena polipeptídica pesada y de la cadena polipeptídica ligera, varían bastante más de un anticuerpo a otro que lo puedan hacer otras secuencias en las regiones variables, con lo que sugieren su responsabilidad en la alta especificidad del lado de unión. Esta hipótesis se apoya en experimentos de marcaje de afinidad, en los que se suministran antígenos sintéticos, conocidos como haptenos, con grupos químicos reactivos que les permiten unirse irreversiblemente al lado de unión de una molécula intacta de anticuerpo. Cuando la molécula se degrada en sus cadenas polipeptídicas componentes, los marcadores de afinidad** *(flechas)* **se encuentran unidos a restos de aminoácidos en las secuencias hipervariables, o en su entorno.**



EL MODELO BASICO DE PLEGADO de la cadena polipeptídica en las zonas constantes *(C)* de la molécula del anticuerpo es de forma cilíndrica y tiene una estructura "en emparedado": la capa superior consta de tres segmentos de cadena polipeptídica y la capa inferior de cuatro *(flechas en color)*, manteniéndose unidas ambas capas por un puente disulfuro (azufre-azufre). Aunque la zona variable, tanto de la cadena ligera como de la pesada, posee un bucle adicional, el modelo global de plegado de la zona constante y de la zona variable guarda un gran parecido, cosa que no se esperaba habida cuenta de las grandes diferencias en sus secuencias de aminoácidos. Los autores proponen que esta estructura zonal común, que ha persistido a pesar de la divergencia evolutiva y de la especialización funcional de las zonas constantes y variables, podría deberse a la presencia, en el interior de ambos tipos, de determinadas zonas de aminoácidos hidrófobos, que producen un modelo muy similar de plegado.



LA CADENA LIGERA de un anticuerpo IgG, procedente de un paciente de mieloma, se pliega en una zona constante *(C)* y en otra variable *(V)*. En el curso de la evolución, las dos zonas han girado una respecto a la otra, de modo que sus capas con cuatro segmentos *(flechas en color)* se orientan en distintas direcciones. Esta rotación ha ido acompañada de cambios en la composición de aminoácidos de las dos zonas, que les permite realizar funciones muy distintas cuando interaccionan en parejas. La asociación de dos cadenas ligeras idénticas en dímeros de Bence-Jones forma una cavidad por donde pueden unirse las moléculas de hapteno. (Las proteínas de Bence-Jones se encontraron en la orina de pacientes de mieloma, y suelen presentarse en parejas o dímeros.) Los dímeros pueden considerarse modelos para un anticuerpo primitivo.

aminoácidos de los 110 que posee la región variable de la cadena ligera y, otros 30, de los 120 aminoácidos de la región variable de la cadena pesada. Debido a la escasa variabilidad en la secuencia de aminoácidos fuera de las regiones hipervariables, muchos investigadores postularon que únicamente las regiones hipervariables se doblarían hacia fuera para formar la superficie, altamente específica, del lado de unión. Se supuso que la función principal del resto de la región variable consistía en presentar las regiones hipervariables en el espacio tridimensional, de tal manera que formaran una cavidad o muesca para que encajara en ella el antígeno.

Las investigaciones de marcaje de afinidad, iniciadas por S. J. Singer y sus colaboradores, en la Universidad de California, en San Diego, corroboraron esa suposición. El marcaje de afinidad, una técnica enzimológica común, implica la unión de marcadores químicos con la cadena polipeptídica de la molécula del anticuerpo en el lado de unión o muy cerca de él. El primer paso de este procedimiento es producir anticuerpos contra el marcador de afinidad. Un antígeno sintético pequeño, o hapteno, verbigracia, un grupo dinitrofenol, se unió a una molécula de proteína portadora y se inyectó en un animal de laboratorio. Se extrajeron después las moléculas del anticuerpo dotadas de una estructura especial, a saber, de complementariedad con el hapteno, toda vez que existe un íntimo acoplamiento entre uno y otras. La combinación normal hapteno-anticuerpo es no covalente y reversible. Para captar el marcador de afinidad, se le agregó al hapteno un grupo químico reactivo, que formará más tarde un enlace covalente con un aminoácido en el lado de unión y así quedará unido de modo irreversible. La molécula de anticuerpo puede, luego, descomponerse en sus cadenas constituyentes, así como también puede determinarse la posición del marcador de afinidad (que es normalmente radiactivo) a lo largo de las cadenas. Cuando se realizaron tales experimentos, los marcadores de afinidad se encontraron generalmente en las regiones hipervariables de las cadenas pesadas y de las cadenas ligeras, lo cual confirmaba la idea de que el lado de unión consta de secuencias hipervariables, aportadas por ambas cadenas polipeptídicas.

Estructuras de los fragmentos de IgG

En los últimos cinco años se han cristalizado algunos componentes de las mo-

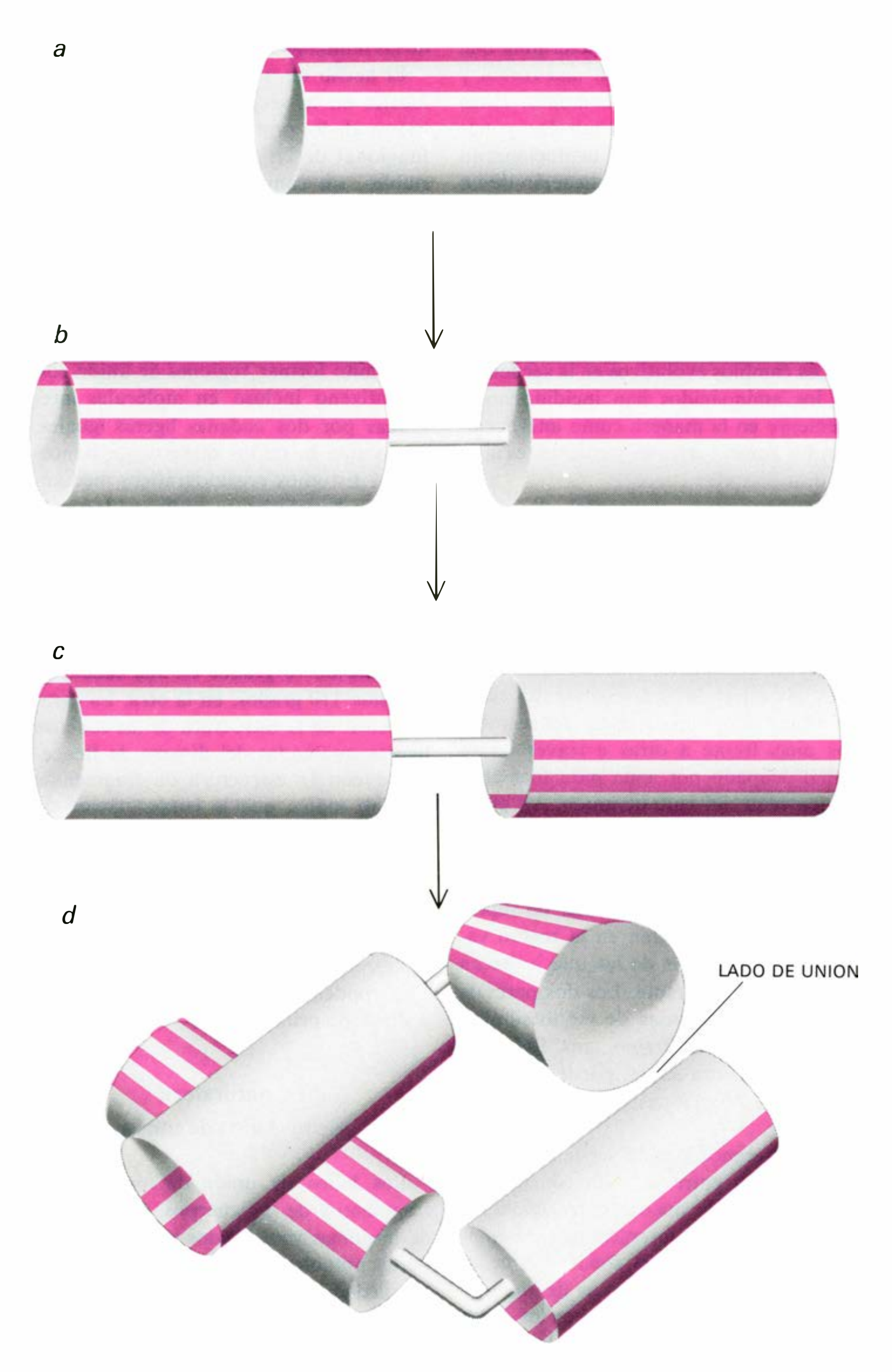
léculas de inmunoglobulina, y así se ha podido examinar su estructura tridimensional por difracción de rayos X. Este trabajo abrió el camino a la comprensión, en términos tridimensionales, de la significación de las secuencias de aminoácidos en el lado de unión. Roberto J. Poljak, Leon M. Amzel y sus colaboradores, de la Johns Hopkins School of Medicine, resolvieron la estructura tridimensional de un fragmento Fab a partir de una proteína de mieloma humano (paciente denominado "New"). David R. Davies, David M. Segal, Eduardo A. Padlan y sus colaboradores, del National Institute of Health, determinaron la estructura de un fragmento Fab de un anticuerpo de ratón con mieloma, consignado con la clave McPC 603. Uno de nosotros (Edmundson), en colaboración con E. E. Abola, K. R. Ely, R. L. Girling y Marianne T. Schiffer, completaron el análisis estructural de un dímero de cadena ligera de Bence-Jones (procedente de un paciente con mieloma denominado "Mcg") en el Argonne National Laboratory. Las tres estructuras anteriores fueron presentadas en el Congreso Internacional de Bioquímica, celebrado en Estocolmo, en 1973. Estos estudios tridimensionales prosiguen, y han hecho posible ya la construcción de modelos detallados de fragmentos Fab y de dímeros de Bence-Jones. Han corroborado la suposición de que antígenos y anticuerpos pueden combinarse química y espacialmente en complejos ordenados.

Los fragmentos Fab y los dímeros de Bence-Jones se parecen en tamaño, forma y características estructurales generales.

Como se había adelantado, a partir de las homologías encontradas en las secuencias de aminoácidos, las proteínas constan de cuatro subunidades, que corresponden a las zonas variables y constantes, aportadas por las cadenas ligera y pesada, en el caso de los fragmentos Fab y, por las dos cadenas ligeras, en el caso del dímero de Bence-Jones. Ambas zonas de cada cadena están unidas covalentemente por tramos extensos de cadena polipeptídica, llamados "regiones de conexión". Las zonas homólogas de las dos cadenas se aparean; cada molécula, en conjunto, forma un tetraedro alargado.

Estructura de las zonas y evolución

Las cuatro zonas tienen una forma cilíndrica y participan de un modelo básico de plegado de cadena polipeptídica llamado plegado de inmunoglobulinas. La estructura de las zonas, en las regiones variables y constantes, posee segmentos rectos de cadena polipeptídica, la mayor parte de los cuales son paralelos al eje más largo de cada subunidad y se hallan dispuestos en dos capas. La primera capa se compone de tres segmentos de cadena polipeptídica y la segunda capa de cuatro segmentos. Los segmentos adyacentes dentro de cada capa se extienden en direcciones opuestas y se estabilizan por enlaces de hidrógeno en-



LA EVOLUCION DEL ANTICUERPO PRIMITIVO empezó, probablemente, con una única cadena polipeptídica de unos 110 aminoácidos, que se plegó en una estructura zonal cilíndrica *(a)*. El gen que codificaba esta cadena se duplicó después, dando origen a dos zonas polipeptídicas idénticas, empalmadas por un corto segmento lineal en cadena *(b)*. La evolución divergente de los genes duplicados cambió gradualmente la composición de los aminoácidos que integraban las dos zonas, diferenciándolas en zonas constante y variable, y determinando que una girase con respecto a la otra al asociarse ambas cadenas completas *(c, d)*.

tre cadenas; adoptan, pues, una forma tridimensional llamada de hoja plegada. Las dos capas rodean un espacio interior compacto de cadenas laterales, hidrófobas, de aminoácidos; las cadenas polipeptídicas se mantienen unidas por un enlace disulfuro. Aunque las regiones variables y constantes difieren sustancialmente en la secuencia de aminoácidos, la presencia de aminoácidos clave en el interior de la estructura determina que las zonas distintas posean una forma y un modelo de plegado comunes. Estas observaciones sugieren que las zonas variables y constantes evolucionaron como resultado de la duplicación de un gene ancestral, después de la cual los genes duplicados divergían hasta cierto punto. Las secuencias de los aminoácidos determinadas por los genes llegaban a ser, por consiguiente, ligeramente distintas.

Los cambios evolutivos en la secuencia de los aminoácidos han incidido notablemente en la manera como interaccionan las zonas variables y constantes apareadas.

En las zonas constantes, las capas de tres segmentos están en el exterior de la molécula, y las capas de cuatro segmentos interaccionan a través de un espacio exento de agua para mantener la ordenación compacta de la región constante. Inversamente, en las zonas variables, las capas de cuatro segmentos están en el exterior, y las capas de tres segmentos unas frente a otras a través de un canal ocupado por agua para formar la concavidad del lado de unión. Así pues, en el curso de la evolución, las dos regiones cilíndricas de cada cadena han girado, una respecto a la otra, de 160 a 170 grados; por consiguiente, las superficies que se enfrentan en la región variable son distintas de las que lo hacen en la región constante. Los dos pares de regiones, constituidas de modo similar, que integran los fragmentos Fab y el dímero de Bence-Jones, pueden realizar, por tanto, funciones enteramente distintas.

Las diferencias en la secuencia de los aminoácidos entre la zona constante y la zona variable pueden correlacionarse con sus distintas interacciones de enlace. Por ejemplo, en la zona variable un modelo alternante de restos de aminoácidos hidrófilos e hidrófobos en las capas de tres cadenas se rompe por la sustitución de aminoácidos hidrófilos por otros hidrófobos en lugares que son importantes para el mantenimiento de la estructura cóncava del lado de unión. De igual modo, la secuencia alternante de aminoácido hidrófilo-aminoácido hidrófobo en las capas de cuatro cadenas de la zona constante se interrumpe por la sustitución de aminoácidos hidrófilos por otros hidrófobos en lugares de íntima interacción entre las dos zonas constantes. Las modificaciones observadas en las secuencias de cada zona se registran en lugares apropiados para su función, pero no menoscaban la integridad de la estructura general del plegado común de las inmunoglobulinas.

El hecho de que unas modificaciones locales en la secuencia de aminoácidos hayan dado origen a la diferenciación funcional de las zonas constantes y variables, sin cambiar su estructura tridimensional global, es clara ilustración de cómo una evolución divergente puede operar a nivel molecular. Efectivamente, la interacción especializada de las zonas apareadas parece ser un rasgo fundamental de la estructura de un anticuerpo, y permite formar regiones de unión con el antígeno incluso en moléculas integradas por dos cadenas ligeras iguales, como sucede en los dímeros de Bence-Jones. Estudios cristalográficos mediante rayos X del dímero de Bence-Jones de Mcg han revelado que las relaciones espaciales entre la zona variable y la zona constante en las dos cadenas ligeras son muy distintas; el ángulo entre los ejes largos de las dos zonas es aproximadamente de 70 grados en una cadena y, de unos 110 grados, en la otra. La comparación de la estructura de las cadenas ligeras apareadas del dímero de Bence-Jones con la estructura de fragmentos de Fab indica que la conformación de una cadena se parece a la de la cadena pesada en fragmentos Fab, mientras que la conformación de la otra cadena se asemeja a la de la cadena ligera de Fab. Puesto que el dímero de Bence–Jones aparece y actúa como un fragmento de Fab, podemos considerarlo modelo de anticuerpo primitivo.

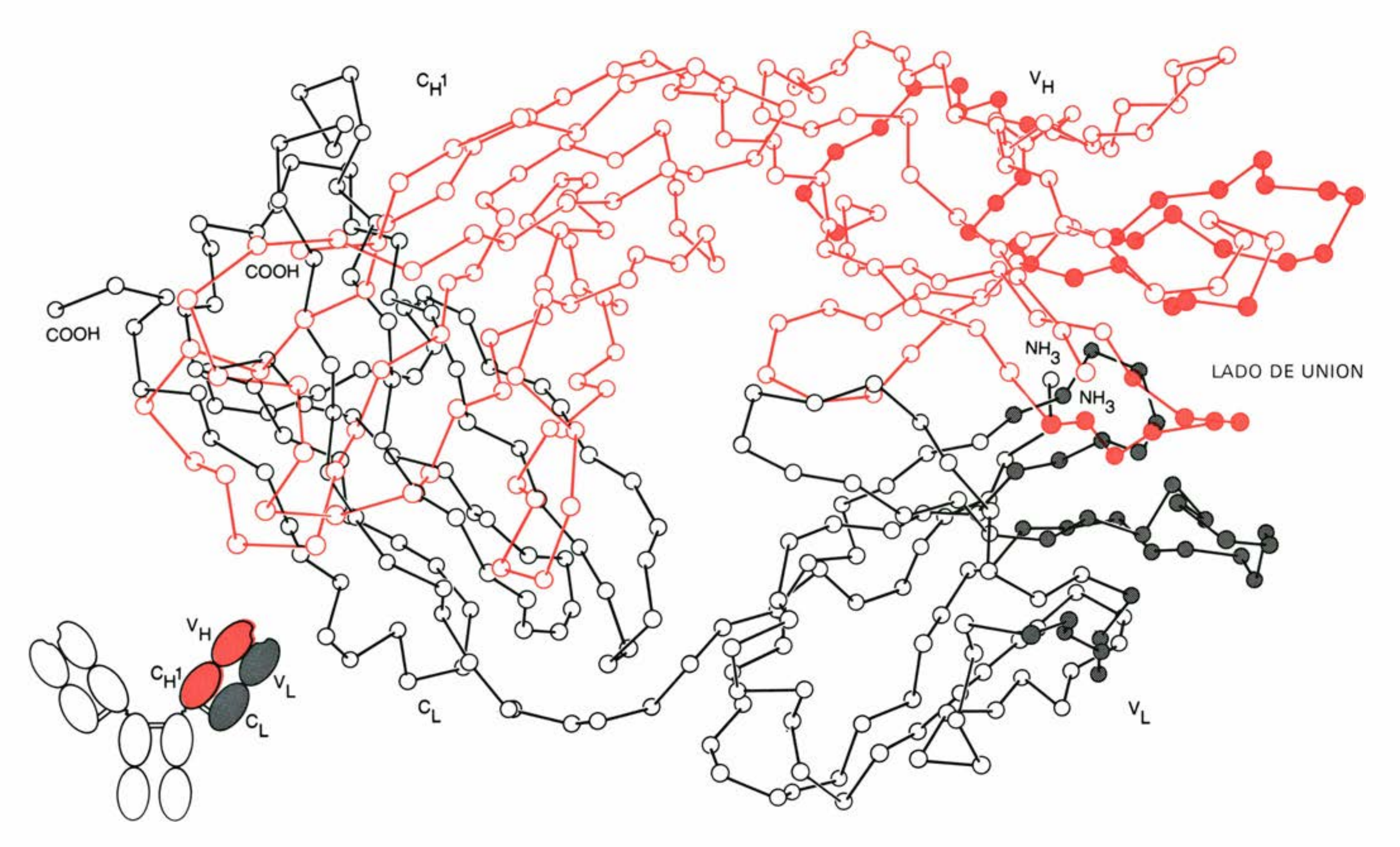
La naturaleza de los lados de unión

Los lados de unión de los distintos fragmentos de anticuerpos, procedentes de mielomas y estudiados hasta ahora, difieren notablemente en tamaño, forma y propiedades de unión con haptenos. Por ejemplo, el fragmento Fab humano del paciente New tiene una cavidad poco profunda que liga una molécula de vitamina K_1 hidroxilada, mientras que el fragmento Fab del ratón McPC 603 posee una hendidura más profunda por donde se une el hapteno fosforilcolina. El dímero de Bence-Jones de Mcg presenta una cavidad cónica todavía más profunda y ancha, conectada a lo largo de un espacio a 17 angstroms de la entrada de la cavidad principal. Cuando se comparan, por medio de difracción de rayos X, las muestras cristalinas de esta proteína con muestras similares en las cuales se ha fijado un dinitrofenilo u otros compuestos del tipo de los haptenos, se hacen evidentes tres lados distintos de unión, dos en la cavidad principal y el tercero en el espacio adjunto. Dependiendo del tamaño y de la estructura química de las moléculas del tipo de los haptenos, los lados pueden ser ocupados uno a uno o los dos a la vez (en tándem). Por ejemplo, compuestos con dos anillos dinitrofenilo adecuadamente distanciados uno de otro, pueden tender un puente con las dos regiones de unión situando un anillo en cada lado.

En cada caso, el lado de unión está formado enteramente por bucles de secuencias hipervariables de aminoácidos fijados a la estructura rígida y constante de las zonas variables. Los bucles hipervariables que forman la superficie del lado de unión en los fragmentos Fab no son grandes: constan como máximo de 17 restos de aminoácidos. En el dímero de Bence-Jones, las regiones de unión que proceden de los bucles hipervariables continúan en las partes más constantes de las capas de tres segmentos. Las seis regiones hipervariables contribuyen todas a la formación de la principal cavidad de unión del dímero de Bence-Jones, pero sólo cinco de estas regiones se hallan implicadas en los fragmentos Fab: las regiones hipervariables primera, segunda y cuarta de la cadena pesada y las regiones hipervariables primera y tercera de la cadena ligera. Unos modelos únicos de cadenas laterales de aminoácidos coinciden en estos bucles para formar una superficie continua, capaz de constituir la pieza complementaria de un antígeno dado. El lado de unión de una molécula de inmunoglobulina resulta, pues, análogo al molde de una máquina impresora en cuanto que puede diferenciarse de un anticuerpo al siguiente sin ningún cambio en la estructura global de la molécula.

Factores que determinan la especificidad

¿Cuáles son los factores que determinan la alta especificidad de un lado de unión para un hapteno o un antígeno? En primer lugar, la forma del lado es complementaria de la forma del antígeno; es cóncava para así poder exponer

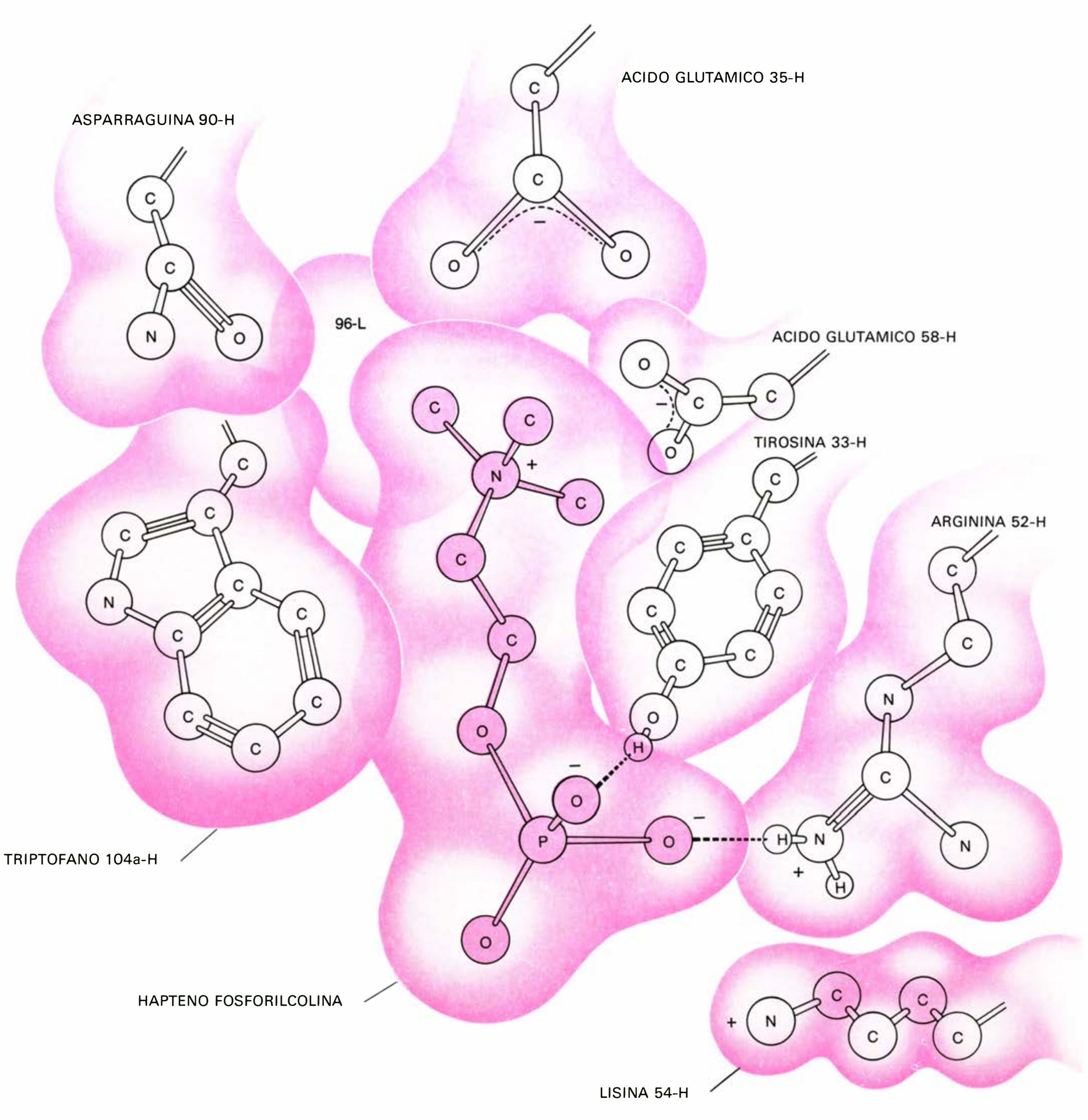


EL ESQUELETO CARBONADO DE UN FRAGMENTO Fab (un brazo de la molécula del anticuerpo), procedente de una proteína de mieloma de ratón denominado McPC 603, muestra como se divide un fragmento en regiones constantes y variables, densamente empaquetadas; éstas constan de zonas apareadas de la cadena pesada *(color)* y de la cadena ligera *(negro)* y unidas por los segmentos "regiones de conexión". El lado de unión en el extremo del fragmento se compone tan sólo de bucles adyacentes de unidades hipervariables *(esferas rellenas, rojas y grises).*

LA ESTRUCTURA DE UN DIMERO DE BENCE-JONES se asemeja sorprendentemente al fragmento Fab. Uno de los dos monómeros de cadena ligera *(color)* del dímero adopta una conformación como la de una cadena pesada de Fab; el segundo se parece a la cadena ligera de Fab.

una gran superficie a la invasión del antígeno. En segundo lugar, las cadenas laterales de los aminoácidos dentro de la cavidad están colocadas con la mayor precisión en orden a aprovechar al máximo las interacciones electrostáticas, de enlaces de hidrógeno y de van der Waals, todas las cuales son más débiles que las interacciones covalentes. (En este sentido, las fuerzas de unión entre una molécula del anticuerpo y otra del antígeno se asemejan a las que existen entre una molécula de enzima y el substrato sobre el cual actúa.)

Por ejemplo, el lado de unión para el hapteno fosforilcolina, en el fragmento Fab del ratón McPC 603, es una cavidad en forma de cuña, cuyas paredes están revestidas de unidades de aminoácidos hipervariables. La fosforilcolina ocupa solamente una pequeña sección de la cavidad; se une asimétricamente, situándose más cerca de la cadena pesada que de la ligera. El grupo fosfato (PO_4) del hapteno, cargado negativamente, está dentro de la distancia del enlace de hidrógeno del grupo hidroxilo (OH) en el lado de unión. Se halla también muy próximo del grupo guanidinio de un resto de arginina cargado positivamente,



EL MODELO DE UNION entre el hapteno fosforilcolina *(esferas en color)* y el lado de unión del anticuerpo McPC 603 ilustra como la forma del lado es complementaria a la forma del hapteno. Además, las cadenas laterales de los aminoácidos que recubren la cavidad se unen al hapteno a través de interacciones débiles, no covalentes, de los tipos: electrostático (atracción de átomos positivos y negativos), enlace de hidrógeno (atracción de átomos de hidrógeno y oxígeno de polarización opuesta) y van der Waals (atracción entre electrones y núcleos atómicos).

proximidad ésta que provoca una fuerte interacción electrostática. El grupo colina del hapteno también parece interaccionar con bucles hipervariables de las cadenas pesada y ligera. El nitrógeno del grupo colina, cargado positivamente, está cerca de la cadena lateral de un resto de ácido glutámico con carga negativa y existen fuerzas de van der Waals entre el grupo colina y los átomos de carbono de las cadenas pesada y ligera.

Se ha logrado establecer estas relaciones mediante análisis por difracción de rayos X de cristales de proteínas. Hay que tener presente que la unión de una pequeña molécula de hapteno a una parte del lado de unión de un anticuerpo en forma cristalina constituye una situación enormemente artificial; *in vivo,* el antígeno ocupa probablemente todo el lado de unión. Por otra parte, los bucles hipervariables, particularmente los que salen hacia el agua que envuelve la molécula del anticuerpo, parecen flexibles; por tanto, el lado de unión podría ensancharse a fin de abrazar un área mayor de la superficie del antígeno. Los estudios cristalográficos del dímero de Bence-Jones indican que las superficies inmediatamente circundantes de la cavidad de unión, pero que no forman parte del interior de la misma, pueden acercarse a las grandes moléculas de un modo lo suficientemente ajustado como para que incluso les permita más interacciones. En los cristales, dichas interacciones mantienen una segunda proteína de Bence-Jones en una orientación específica tal, que una de las cadenas laterales de sus aminoácidos entra en la cavidad del dímero y se sitúa en el lado apropiado para acoplarse al antígeno.

Como ya hemos dicho, la zona constante y la zona variable de cada cadena polipeptídica de los fragmentos Fab se hallan unidas covalentemente a través de tramos de cadena que reciben el nombre de regiones de conexión. En el dímero de Bence-Jones, esta región también está doblada en una extensión mayor en la cadena análoga a la pesada que en la ligera. No obstante, R. Huber y sus colegas alemanes han descrito recientemente una proteína IgG en la cual las regiones de conexión en las partes Fab de la molécula no presentan tan gran diferencia entre la cadena ligera y la cadena pesada.

Del conjunto de estas observaciones se postula que ambas cadenas, pesada y ligera, pueden realizar grandes movimientos de bisagra a nivel de las regiones de conexión, pero es probable que estos movimientos no se den en todos los anticuerpos IgG. Sin embargo, la capacidad para cambios de conformación en dichos segmentos conectores podría ser muy importante en el mecanismo por el cual la unión de un antígeno con ciertos anticuerpos da origen a una variedad de respuestas inmunológicas, entre las que se incluyen la disolución de células sanguíneas por un factor sanguíneo conocido como complemento, la liberación de histamina a partir de las células cebadas y la activación de la síntesis de los mismos anticuerpos por las células plasmáticas.

La modificación de la especificidad

El estudio de finos modelos estructurales, construidos a partir del análisis por difracción de rayos X, ha contribuido notablemente al esbozo teórico de como la geometría del lado de unión de un anticuerpo, y, consecuentemente, su especificidad por un antígeno, podría modificarse efectuando cambios en la secuencia de aminóacidos de las regiones hipervariables. Por ejemplo, la simple sustitución de una unidad de ácido glutámico cargada negativamente por otra de arginina con carga positiva, en una cadena polipeptídica del lado de unión, no afectaría probablemente a la topología global de la superficie hipervariable, pero cambiaría profundamente sus propiedades de unión con antígenos. Además, inserciones y supresiones dentro de los bucles hipervariables alterarían su tipo de plegado y podrían cambiar la forma y tamaño de la cavidad hipervariable. Finalmente, existe la posibilidad de que las sustituciones, las supresiones o las inserciones de aminóacidos en las partes más profundas de la interfase entre las partes variables alteren las interacciones entre estas dos partes, dando como resultado un cambio en las posiciones relativas de los bucles hipervariables en la cadena pesada y en la cadena ligera y originando cambios de especificidad significativos.

Recapitulación

En resumen, los impresionantes progresos de la inmunoquímica en los últimos años han trazado una imagen mucho más clara de como una molécula de anticuerpo se complementa con un antígeno y han aportado mayor información sobre la evolución de la estructura del anticuerpo. Ahora, parece estar al alcance de la mano una explicación molecular racional de la capacidad del sistema inmune del cuerpo para responder a una variedad aparentemente infinita de estímulos antigénicos.

La percepción de objetos en movimiento

En el sistema visual humano existen circuitos nerviosos especializados que separan la información referente a la dirección en que un objeto se está moviendo de los detalles referentes a la forma del mismo

Robert Sekuler y Eugene Levinson

La capacidad de dar una respuesta rápida y adecuada a un objeto en movimiento es, con frecuencia, cuestión de vida o muerte. No es de extrañar, por tanto, que el sistema visual de los mamíferos haya evolucionado hacia la obtención de unos mecanismos nerviosos altamente especializados para llevar a cabo esta tarea. Los investigadores han obtenido abundante información respecto a la forma en que funcionan; en diferentes especies, estos circuitos nerviosos complejos, a través de unos experimentos fisiológicos que consisten en la implantación de microelectrodos en la corteza visual del cerebro de un animal. Sin embargo, es obvio que el estudio de esa misma función en el hombre requiere la utilización de técnicas menos intrusivas. Aunque los procedimientos desarrollados en este sentido sean forzosamente indirectos, han proporcionado una información sorprendente sobre los mecanismos que utiliza el sistema visual humano para percibir objetos en movimiento.

Lo primero que se necesita en cualquier procedimiento psicofísico de este tipo es la presentación de estímulos específicos. En tal caso, no sirven de mucho las cosas que vemos en movimiento en la vida diaria, pues su repercusión sobre el sistema visual es, a la vez, demasiado complicado y demasiado débil para poder obtener mediciones fiables. Por tanto, aunque la finalidad última sea entender por qué los objetos en movimiento se perciben tal y como se nos presentan en la vida diaria, se debe empezar trabajando con "imágenes-objeto" en movimiento que sean, a la vez, menos complicadas y más potentes que los objetos ordinarios.

A principios del siglo XIX, los investigadores en el campo de la percepción visual humana, se dieron cuenta ya de que una exposición prolongada a un estímulo sencillo e intenso podía poner de manifiesto propiedades del sistema de percepción del movimiento que no podían observarse con otros estímulos más variables. Así, el investigador británico R. Addams observó que una fijación visual prolongada sobre un movimiento en una dirección provocaba a menudo un postefecto singular. Addams fijó la vista, durante unos segundos, en una cascada cerca de Loch Ness. Luego, cuando desvió la mirada hacia las rocas vecinas, éstas le parecieron moverse hacia arriba. El fisiólogo checo Jan Purkinje había ya experimentado anteriormente una ilusión similar tras contemplar un desfile de jinetes, durante una hora aproximadamente.

En nuestro laboratorio de la Northwestern University, conseguimos efectos parecidos mediante estímulos generados electrónicamente. Descubrimos un tipo de estímulo que sirve para obtener dichos efectos: consiste en un dibujo patrón en movimiento, constituido por puntitos brillantes "pintados" con la ayuda de una computadora sobre la pantalla de un tubo de rayos catódicos. Normalmente, los puntos se hallan distribuidos al azar en la pantalla. Los extremos laterales de la pantalla están conectados funcionalmente: cuando un punto desaparece por un lado de la pantalla, reaparece inmediatamente por el otro. Desde el punto de vista del observador, es como si mirara una lámina infinitamente grande de puntos en movimiento, a través de una abertura estacionaria. Al ser aleatoria la distribución de los puntos en el área, nos permite estudiar el impacto visual que provoca su movimiento, eliminando los complicados efectos debidos a la percepción de los contornos.

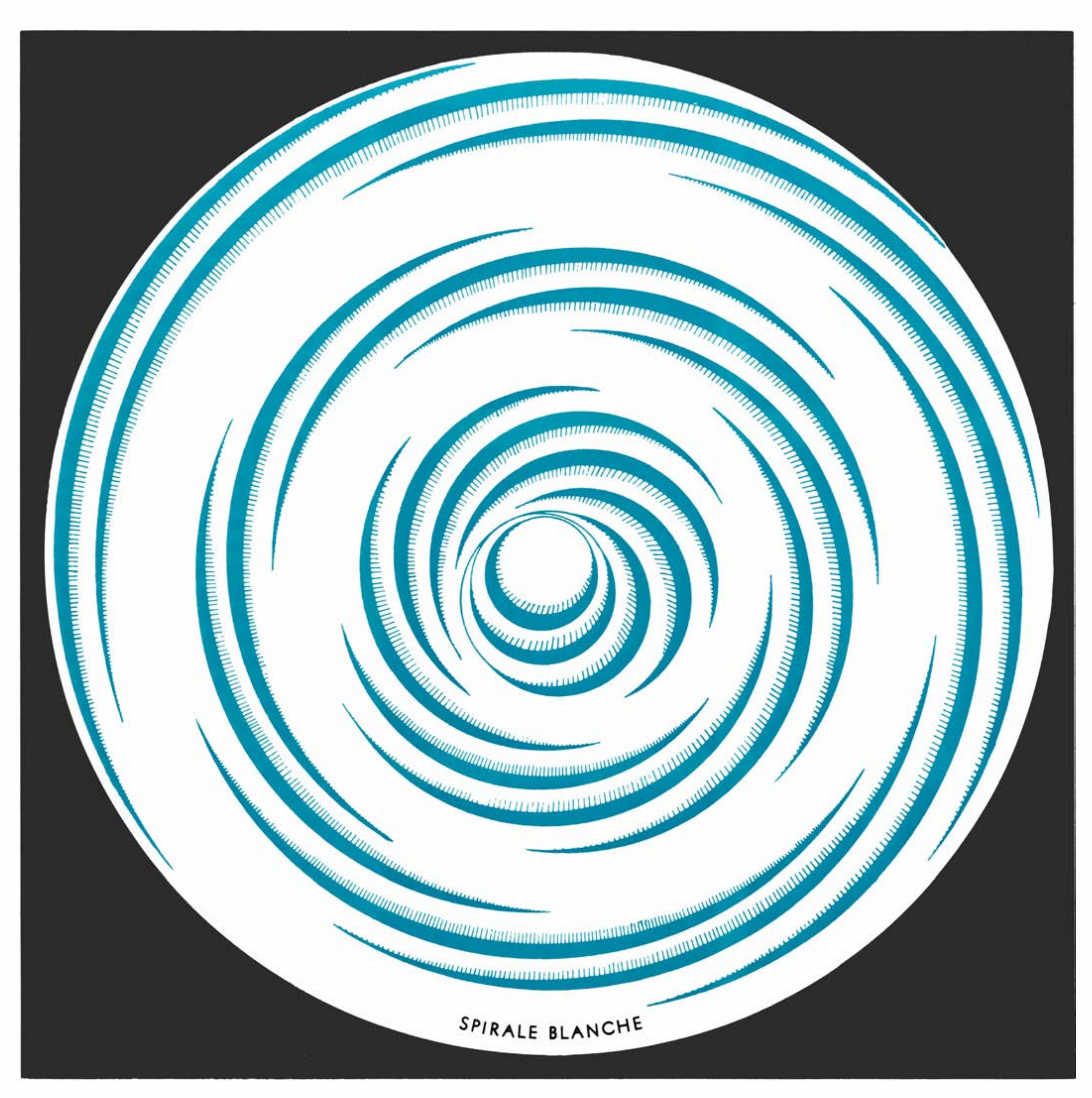
Un dibujo patrón de franjas paralelas, claras y oscuras, alternativamente, constituye otro tipo de "imagen-objeto", formada en un tubo de rayos catódicos al igual que el tipo anterior. Los dibujos de esta clase se denominan enrejados, y se les puede dar un movimiento en dirección perpendicular a las franjas, a cualquier velocidad. Tanto con los dibujos de enrejados como con los de puntos, se puede generar un movimiento continuo, repetitivo y cuya duración sea la que se desee. Se puede así permitir la acumulación del efecto momentáneo y aislado del movimiento, produciendo una repercusión total cuya intensidad sea suficiente como para obtener medidas fiables.

La respuesta del sistema visual a cualquier "imagen-objeto", incluidas las que están en movimiento, depende del contraste entre la parte oscura y la parte clara de la imagen. El hecho de que los televisores tengan mecanismos de ajuste diferentes para el contraste y para el brillo sirve para recordar que son dos variables distintas. Para cualquier nivel de brillo medio, se puede ajustar el contraste de forma independiente, volviéndose la imagen más o menos borrosa. Una forma de medir la eficacia con que una determinada imagen de televisión estimula al sistema visual consiste en ajustar el contraste hasta el punto en que el observador pueda apenas ver que la pantalla de televisión no presenta una luminosidad perfectamente uniforme. Si la imagen concreta que está en la pantalla resulta ser una imagen a la que el observador es muy sensible, permanecerá visible aun cuando el contraste sea muy bajo.

De lo anterior se deduce que se puede

usar el umbral de contraste, en el cual desaparece una "imagen-objeto", como índice de la sensibilidad del observador a ese objeto particular. Un bajo umbral de contraste significaría una elevada sensibilidad, y viceversa. Se puede estudiar de la misma manera los sistemas de enrejados o de puntos, generados electrónicamente; al disminuir el contraste hasta el punto en que el observador apenas distinga el dibujo patrón, se puede obtener una medida del umbral de contraste del observador para ese dibujo.

Ulker T. Keesey, de la Universidad de Wisconsin, y Janos Kulikowski, de la Universidad de Manchester, han demostrado que los observadores poseen, no sólo uno, sino dos umbrales de contraste diferentes para muchas "imágenes-objeto" en movimiento. Supongamos, por ejemplo, que estamos trabajando con un enrejado de baja frecuencia (o sea, con franjas anchas) que se mueve a través de la pantalla de un tubo de rayos catódicos. Se empieza poniendo el contraste a cero; luego, se aumenta gradualmente hasta el punto preciso en que el observador detecta que hay algo presente en la pantalla. En realidad, el observador no ve el enrejado; lo que percibe es



LA ESPIRAL BLANCA pertenece a un conjunto de 12 ilusiones ópticas que aparecen con un movimiento rotatorio y que se denominan Rotorrelieves, designación que les dio el artista Marcel Duchamp en 1935. Esta espiral pone de manifiesto una importante propiedad del sistema visual humano: la independencia que presentan entre sí las vías nerviosas responsables de la percepción de direcciones de movimiento opuestas. Si se hace girar una copia de este dibujo en el sentido de las agujas del reloj, a una velocidad de 33 revoluciones por minuto aproximadamente (por ejemplo, sobre un plato de tocadiscos), cualquier observador debe ser capaz de distinguir dos espirales superpuestas: una espiral obscura que posee franjas estrechas y numerosas y parece presentar un movimiento dirigido hacia el centro de rotación, y una espiral blanca, con franjas más anchas y menos numerosas, que parece moverse hacia la periferia, alejándose del centro. En realidad, ambas configuraciones espirales entretejidas tienen una posición fija la una respecto de la otra; no obstante, cuando están girando y se las observa, (de preferencia con un solo ojo), ambas direcciones de movimiento se separan claramente. Reproducción de la colección de A. d'Harnoncourt.

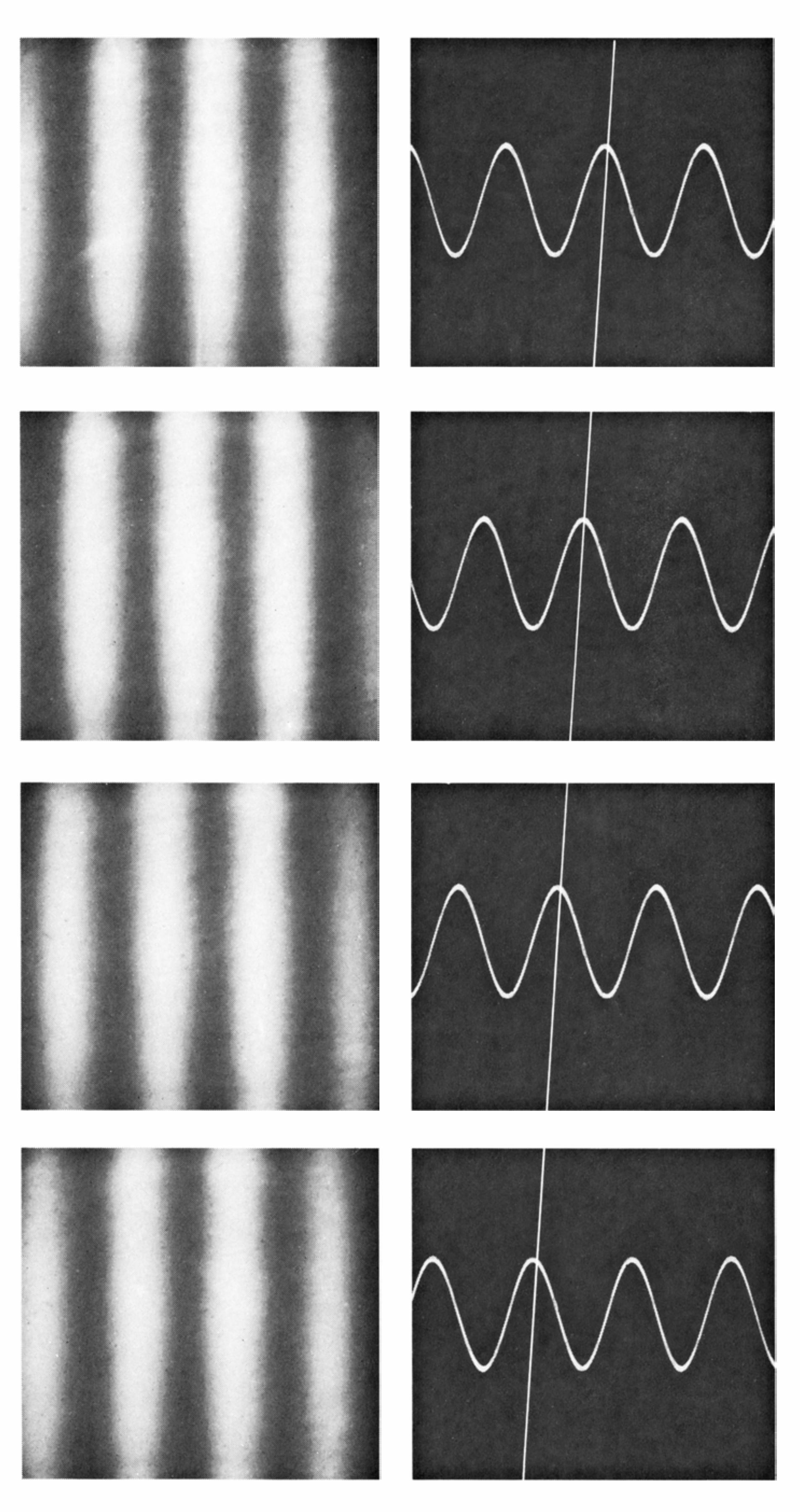


FIGURA EN MOVIMIENTO, compuesta por franjas paralelas alternativamente claras y obscuras, cuya aparición puede provocarse en la pantalla de un tubo de rayos catódicos y utilizarla como una "imagen-objeto" apta para experimentos psicológicos con observadores humanos. En la secuencia de fotografías de la izquierda, realizada en el laboratorio de los autores, se muestra un enrejado que se mueve hacia la izquierda, atravesando la pantalla a una velocidad constante. La curva sinusoidal que acompaña a cada fotografía es una representación de la intensidad luminosa a lo ancho de la pantalla en el momento de cada exposición. La larga línea oblicua une crestas de curva en intervalos sucesivos (*a la derecha*).

una procesión de ondas informes. Aunque puede distinguir la dirección en que se mueven las ondas, las franjas mismas del enrejado permanecen indistintas. Si se sigue aumentando el contraste, se puede llegar a ver las franjas. Resulta evidente que, con un contraste muy bajo, se puede ver el movimiento de una "imagen-objeto" aunque no las características de su forma, mientras que, con un contraste mayor, se puede discernir tanto la forma como su movimiento. La existencia de umbrales diferentes para la percepción del movimiento y para la percepción de la forma lleva, como es natural, a la sospecha de que existen dos vías ópticas diferentes en el sistema nervioso humano, a las que compete analizar estos dos aspectos de una "imagen-objeto" en movimiento.

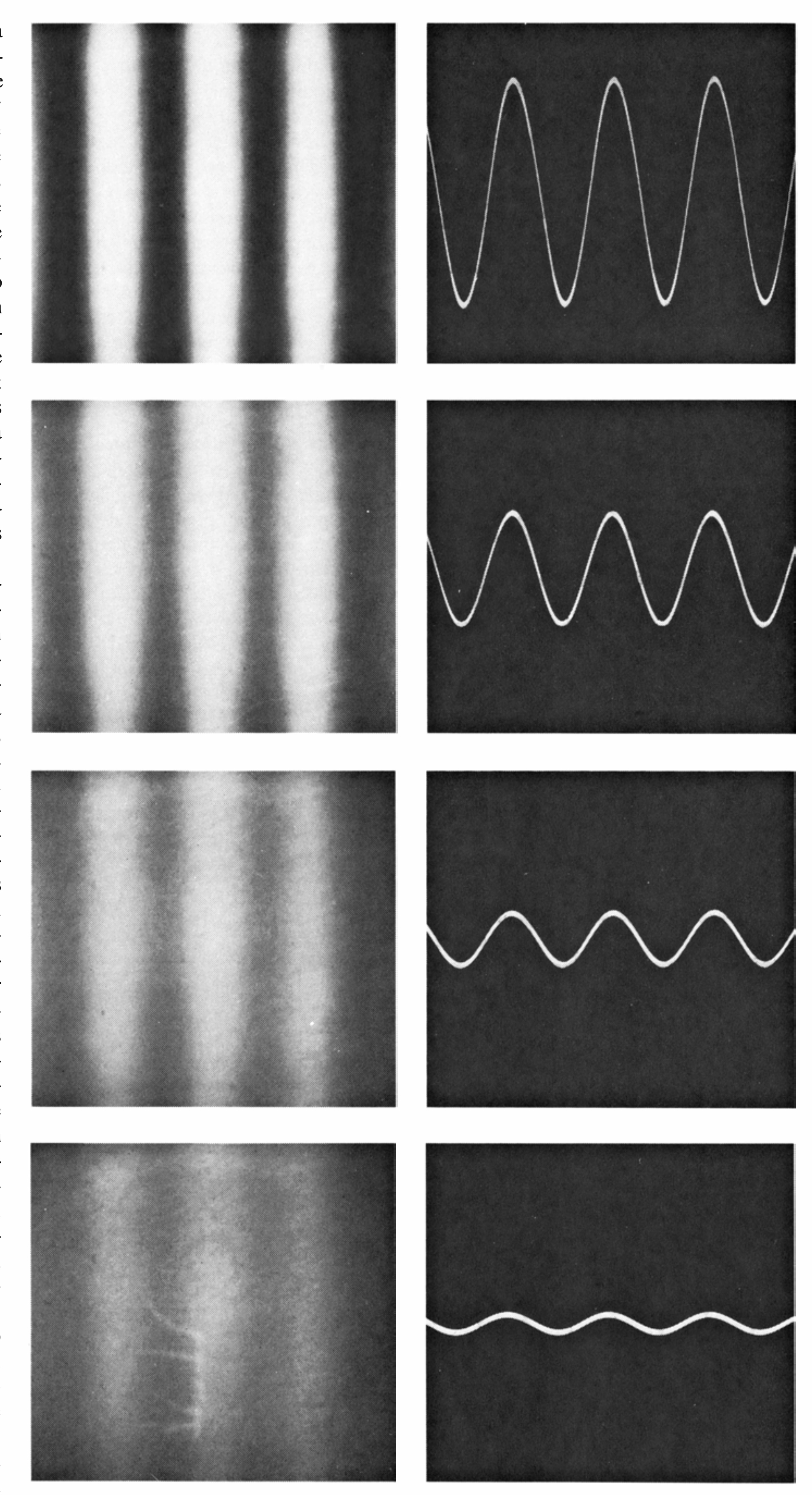
Es evidente que la información que nos puede proporcionar la observación de "imágenes-objeto" en movimiento sobre la reacción del sistema nervioso al movimiento tiene un límite. Se necesitan más métodos analíticos. Uno de ellos es el procedimiento psicofísico llamado adaptación selectiva, que suele utilizarse para estudiar el olfato, la audición y el gusto y, también, la visión. Esta técnica permite "disecar" cualquier sistema sensorial sin dolor alguno. Lo que se hace en la adaptación selectiva es medir la detección de varios estímulos comparando las respuestas dadas antes y después de someter el sistema sensorial a un estímulo especial, de gran intensidad, denominado estímulo adaptador. Pongamos por caso que la exposición a un estímulo adaptador dificulte la detección de un estímulo de prueba, pero tenga un efecto escaso o nulo sobre la capacidad para detectar otro estímulo de prueba distinto. Sería lógico concluir que los mecanismos necesarios para la detección del primer estímulo de prueba coinciden en parte con los mecanismos afectados por el estímulo adaptador, pero no así los mecanismos necesarios para la detección del segundo estímulo de prueba. Generalmente, se piensa que el estímulo adaptador produce fatiga, es decir reduce la sensibilidad de unos determinados mecanismos del sistema nervioso; esta mengua de la sensibilidad hace más difícil la detección de estímulos que son procesados por el mecanismo afectado.

Hace aproximadamente doce años, unos experimentos de adaptación selectiva establecieron que el sistema visual humano empleaba mecanismos selectivos para la dirección, en alguna etapa de su proceso nervioso. Leo Ganz, que ahora está en la Universitat de Stanford, y uno de nosotros (Sekuler) sometieron a unos observadores a la visión de enre-

jados verticales en movimiento hacia la derecha o hacia la izquierda. Los estímulos adaptadores poseían un contraste elevado; su misión era desensibilizar selectivamente los mecanismos que responden a una u otra dirección de movimiento, si es que existían dichos mecanismos. Tras un corto período de adaptación, se medían los umbrales de contraste para un enrejado cuyo movimiento coincidía con la dirección del estímulo adaptador o bien iba en la dirección opuesta. Había una diferencia significativa en la visibilidad del estímulo de prueba, según en qué dirección se movía: tras la adaptación, se hacía mucho más difícil ver el enrejado que se movía en la misma dirección que el enrejado adaptador. Desde entonces, este efecto –llamado adaptación específica a la dirección– ha sido reproducido en numerosas ocasiones.

Naturalmente nos preguntamos si estos mecanismos –objeto de nuestros experimentos– estaban relacionados con ciertas neuronas que presentaran fuertes preferencias para la dirección del movimiento. Una neurona selectiva para la dirección se dispara fuertemente cuando, a través de su campo visual, pasa, en una dirección, un estímulo de orientación adecuada; en cambio, su ritmo de producción de impulsos queda muy reducido si el mismo estímulo atraviesa el campo en la dirección opuesta. Estas células específicas para la dirección abundan en la corteza visual del mono y del gato. Este tipo de neuronas podría ser responsable de la capacidad del hombre para ver objetos en movimiento, ya que son capaces de distinguir entre direcciones opuestas. Aunque no se sepa con seguridad si el cerebro humano posee exactamente los mismos tipos de células que el cerebro del mono, se sabe que, en áreas de la visión que no están en relación directa con el movimiento, las capacidades visuales del hombre son casi idénticas a las de algunos simios. Además, parece ser que éstos experimentan postefectos cinéticos de forma muy similar a los seres humanos.

Varios investigadores han intentado determinar si la actividad eléctrica del cerebro humano muestra también selectividad para la dirección. Puede aplicarse unos electrodos al cuero cabelludo de un observador y luego medir la respuesta másica del tejido cerebral subyacente a los electrodos, mediante una amplificación y un proceso de señales adecuados. Con este método, se puede llevar a cabo el experimento de adaptación específica a la dirección, midiendo la respuesta cerebral másica. Así, un investigador midió la respuesta cerebral a un dibujo patrón



EL UMBRAL DE CONTRASTE es el nivel de contraste que existe entre las partes obscura y clara de una "imagen-objeto", y, en el cual, apenas se puede ver la imagen. Dicho umbral sirve como un índice de la sensibilidad de un observador a esa imagen. Por ejemplo, en la secuencia de fotografías de la izquierda, el contraste de los cuatro enrejados estacionarios –que en sus demás características son idénticos– disminuye progresivamente de arriba abajo. Se ha comprobado que los observadores humanos tienen dos umbrales de contraste diferentes frente a "imágenes-objeto" en movimiento: un umbral para percibir el movimiento y otro para la forma. Las curvas de la derecha representan la intensidad luminosa.

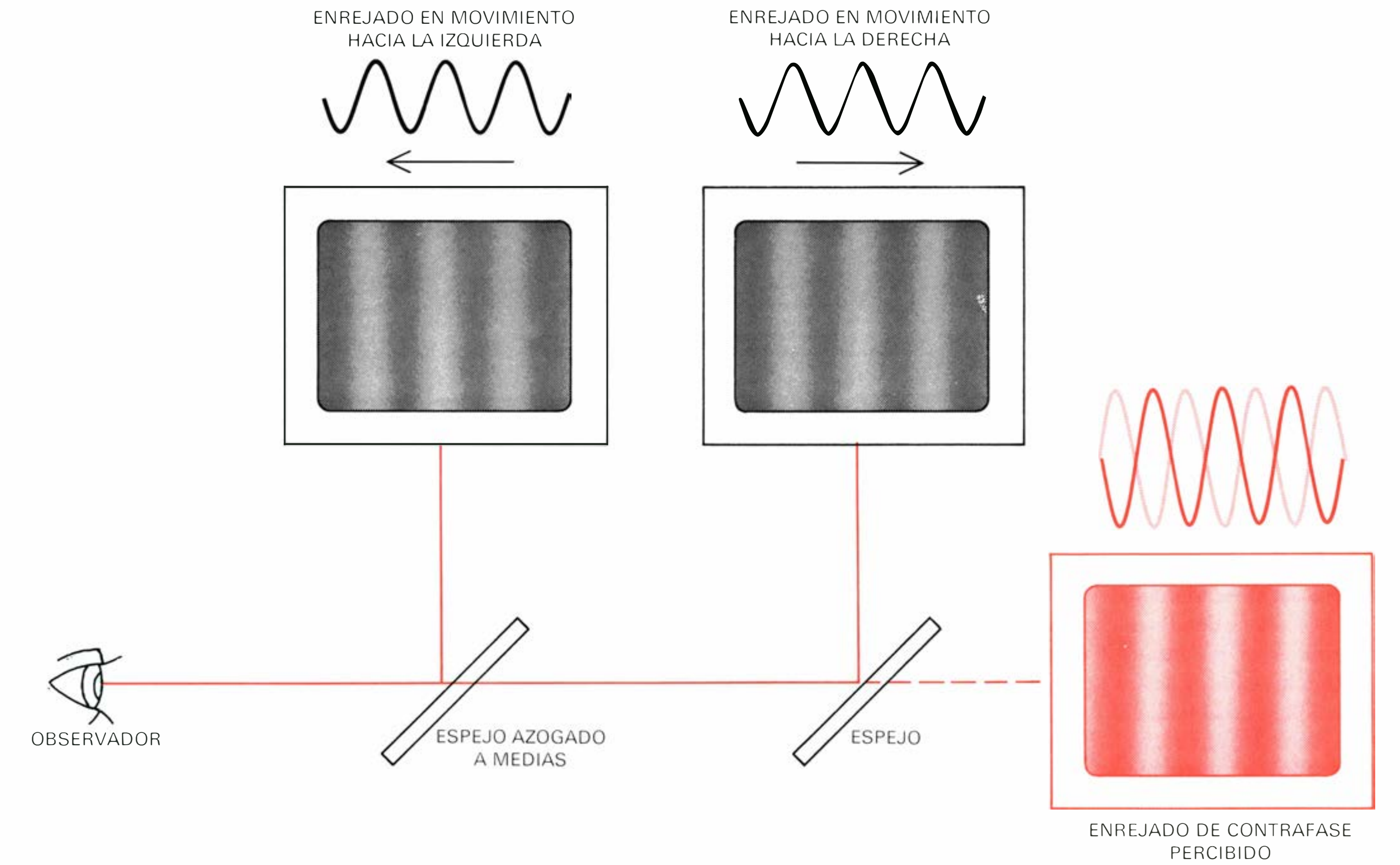
que se movía primero en una dirección y luego en la dirección opuesta. La adaptación introdujo un cambio en la respuesta cerebral al movimiento cuya dirección coincidía con la del dibujo adaptador, pero no quedó alterada la respuesta al movimiento en la dirección opuesta.

En resumen, los experimentos de adaptación selectiva proporcionan pruebas de que el sistema visual humano comprende algunos mecanismos selectivos para la dirección. Sin embargo, la selectividad direccional, que se demuestra con las técnicas de adaptación, no sacia nuestra curiosidad acerca de la forma en que el sistema responde a un estímulo en movimiento. Por ejemplo, el tipo de selectividad observado en estos experimentos no implica forzosamente que haya detectores diferentes para las distintas direcciones del movimiento. La adaptación específica a la dirección podría reflejar una actividad en neuronas selectivas para esta característica del movimiento, de donde la información pasaría a acumularse posteriormente en una etapa de detección, que no sería específica para la dirección. Los experimentos de adaptación confirman que el sistema visual analiza la dirección del movimiento, pero quisiéramos saber como es procesada la información después de este análisis preliminar.

Siguiendo una metodología propia de los ingenieros, recurrimos a dos tubos de rayos catódicos diferentes y en uno generamos un enrejado que se movía hacia la derecha, mientras que en el otro produjimos un enrejado similar que iba hacia la izquierda. De las imágenes de ambos tubos conseguimos una combinación óptica (mediante un espejo azogado a medias) que el observador puede mirar, ajustando el contraste del enrejado combinado con el nivel umbral.

A pesar de las numerosas mediciones que hemos hecho con la ayuda de este procedimiento básico, hay una condición especial que presenta un interés particular: es el caso en que los enrejados posean franjas de la misma anchura y contraste, deriven a la misma velocidad, pero se muevan en direcciones opuestas. En este caso concreto, el estímulo combinado no presenta ninguna dirección de deriva: ¡lo único que hace es invertir su fase repetidamente! En cualquier punto de la retina del observador, el enrejado combinado va cambiando de claro a obscuro y al revés, sin que se observe ningún movimiento evidente en ninguna dirección. De hecho, cuando un observador echa una ojeada rápida a un enrejado combinado de este tipo, no ve movimiento; sólo parece que el enrejado presenta pulsaciones, en que las franjas obscuras se hacen claras y viceversa.

Ahora supongamos que el observador mira uno solo de los enrejados en movimiento que componen la combinación y ajusta el contraste hasta el punto en que apenas ve algo. Repite luego el mismo proceso con el otro enrejado componente, también presentado aisladamente. Ambos umbrales aparecerán en un valor de contraste muy similar. Posteriormente, el observador mira el enrejado combinado, manteniendo el contraste de cada componente en su propio valor umbral. Suponiendo que los enrejados que se mueven en direcciones opuestas se detectaran independientemente, el enrejado combinado también se encontraría en el umbral de contraste. Por otra parte, el enrejado combinado tan sólo se podría ver con claridad si cada enrejado componente presentara un contraste suficiente para verse ais-

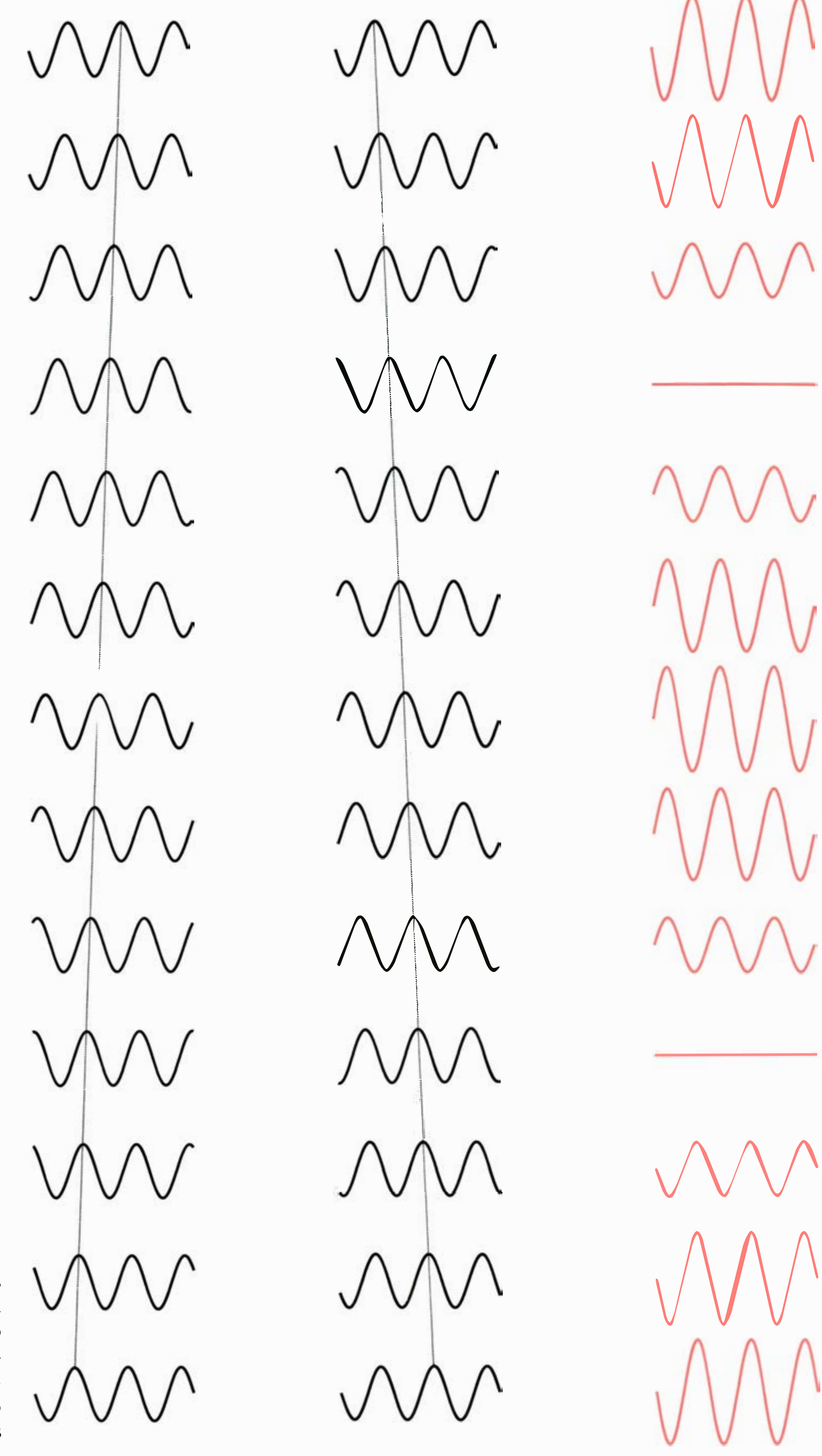


EL ENREJADO DE CONTRAFASE, que también se conoce bajo la denominación de enrejado de fase alternante, se logra formar combinando dos enrejados que se muevan en direcciones opuestas y que tengan igual contraste e igual anchura de las franjas. En el montaje experimental que aquí se presenta, se usan dos tubos de rayos catódicos diferentes; en uno, el enrejado se mueve hacia la derecha mientras que, en el otro, el enrejado se mueve hacia la izquierda. Las imágenes de ambos tubos se combinan ópticamente mediante un espejo azogado a medias y se presentan como estímulo al observador. El estímulo visual resultante *(en color)* no presenta ninguna dirección de movimiento; sólo invierte la fase repetidamente. El ojo sí puede descomponer fácilmente la imagen intermitente en sus imágenes componentes cuyo movimiento es opuesto.

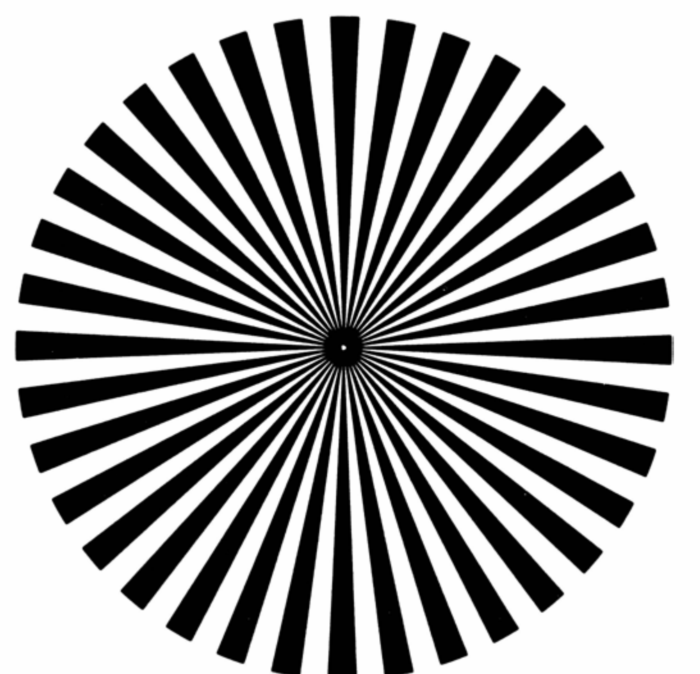
lado. El aspecto que de hecho presenta el enrejado combinado con el nivel umbral sugiere que este razonamiento es correcto: cuando la combinación se halla justamente en el umbral, el observador ve primero un componente y luego otro. La detección del estímulo combinado depende, al parecer, de dos mecanismos visuales independientes, responsables cada uno de la detección de un solo componente de dirección; estos mecanismos tienen una sensibilidad que va fluctuando ligeramente con el tiempo.

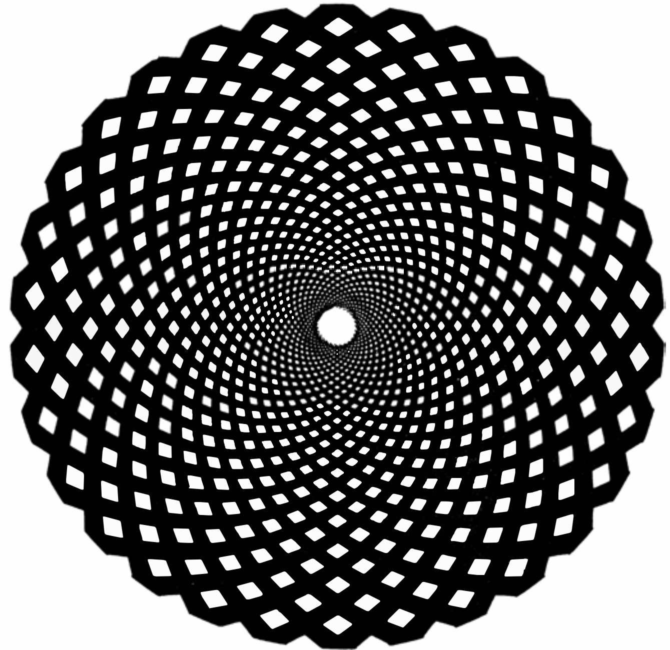
La posibilidad de que los componentes de un enrejado combinado, que se mueven en direcciones opuestas, sean detectados por mecanismos sensibles a la dirección e independientes, nos lleva a una hipótesis interesante: el umbral de contraste para la detección de un enrejado combinado debería ser exactamente el doble del umbral de contraste para la detección de uno de sus componentes presentado aisladamente. Es de esperar esta doble magnitud, pues, como ya hemos explicado, si cada componente del enrejado combinado se halla en su umbral de contraste, ambos componentes están en el umbral y el enrejado combinado adquiere, en conjunto, un contraste total que es igual a dos unidades de umbral. De acuerdo con esta suposición, en nuestros experimentos hemos establecido comparaciones entre el umbral de contraste de un enrejado combinado y el de cualquiera de sus enrejados componentes en movimiento, tomados aisladamente. Hemos repetido las mediciones para distintas anchuras de las franjas y distintas velocidades de movimiento de los enrejados. Los resultados son inequívocos: para detectar un enrejado combinado se necesita un contraste que sea doble del requerido para detectar uno de los enrejados componentes del estímulo combinado.

Puesto que el enrejado combinado, constituido por componentes de igual contraste que se mueven en direcciones opuestas, tiene una gran importancia en el estudio de la visión, se le ha dado un nombre especial: enrejado de contrafase o de fase alternante. A fin de examinar otras aplicaciones del enrejado de contrafase, en primer lugar podemos relacionar nuestros resultados psicofísicos con algunos estudios recientes sobre neuronas singulares de la corteza visual del gato. John G. Robson y Graham Cooper, de la Universidad de Cambridge, han obtenido un registro de las respuestas de células aisladas a enrejados de movimiento unidireccional y a enrejados de contrafase. Una célula, que



LA DESCOMPOSICION DE UN ENREJADO DE CONTRAFASE en dos enrejados que se mueven en direcciones opuestas es el fenómeno que aquí representamos gráficamente. La primera columna indica las sucesivas curvas de intensidad luminosa para un enrejado sinusoidal que se mueva hacia la izquierda a una velocidad constante. La segunda columna indica las curvas correspondientes a un enrejado que se mueva hacia la derecha a la misma velocidad. Cada una de las curvas de color, que se hallan en la tercera columna, da la suma algebraica de las otras dos curvas de la misma línea; las curvas de esta columna, que representan el enrejado de contrafase, atraviesan un ciclo completo de inversión de fase en los 12 intervalos de tiempo de muestra. La amplitud de la curva superior de la tercera columna, medida de pico a pico, es el doble de la amplitud de cada una de las otras dos.





EL POSTEFECTO NEGATIVO, que se da en la percepción del movimiento, se puede poner de manifiesto mediante estos dos dibujos. Coloque a su izquierda una copia del dibujo radial "adaptador" sobre un plato de tocadiscos. Fije la vista en el centro del dibujo mientras va girando a 33 $^{1}/_{3}$ revoluciones por minuto. Después de ello, mire el centro del dibujo estacionario que se halla a la derecha. Durante breves instantes, el segundo dibujo parecerá estar girando en el sentido opuesto. Los experimentos llevados a cabo con estímulos móviles, que se generan electrónicamente, revelan efectos de adaptación similares *(véase la ilustración inferior de la página siguiente).*

reacciona igualmente bien ante cualquier dirección de la deriva de un enrejado, se dispara con la misma intensidad ante un enrejado de contrafase que ante un enrejado móvil de igual contraste. Sin embargo, una célula que sea selectiva para la dirección reacciona ante un enrejado de contrafase con una intensidad igual a la mitad de la reacción que da ante un enrejado móvil de igual contraste. Por tanto, la célula selectiva para la dirección parece reaccionar tan sólo ante el componente que representa la mitad del contraste del enrejado de contrafase y que deriva en la dirección preferida de la célula. Paralelamente, se hicieron mediciones psicofísicas en los gatos: sobre la base de una técnica conductista, John M. Camisa, Randolph Blake y uno de nosotros (Levinson) descubrieron que el umbral de contraste del gato para un enrejado de contrafase es el doble del requerido para un enrejado de movimiento unidireccional. Estos resultados son similares a nuestras observaciones psicofísicas en sujetos humanos.

La doble diferencia existente entre los umbrales de contraste de los enrejados de contrafase y los de movimiento unidireccional no es el único motivo que tenemos para suponer que el sistema visual humano descompone el enrejado de contrafase en sus componentes móviles aislados. Una adaptación previa a un enrejado que se mueve en una dirección puede alterar radicalmente el aspecto que presenta el enrejado de contrafase. Por ejemplo, si se mira un enrejado en movimiento hacia la izquierda –desensibilizando de esta forma los mecanismos sensibles al movimiento hacia la izquierda– cualquier enrejado de contrafase que se observe seguidamente parecerá, en un principio, moverse hacia la derecha. Pasados varios segundos, el enrejado vuelve a tomar su apariencia pulsátil normal. El tiempo que dura la impresión de que el enrejado de contrafase se mueve hacia la derecha es, presuntamente, el tiempo durante el cual la vía selectiva para el movimiento hacia la izquierda resulta menos sensible que la vía selectiva para el movimiento hacia la derecha. Al desvanecerse el efecto de la adaptación, la sensibilidad de ambas vías vuelve a la igualdad y la apariencia pulsátil reaparece. Desde luego, parece bastante extraordinario que el sistema visual pueda extraer los dos componentes móviles de un estímulo que oculta tan bien el movimiento como lo hace un enrejado de contrafase. Sin embargo, se conoce la existencia de casos similares en otros campos sensoriales. La separación de los componentes móviles recuerda la capacidad del oído para descomponer un acorde en sus tonos armónicos individuales.

La capacidad que posee el sistema visual para utilizar detectores independientes para el análisis de los componentes de una imagen compleja en movimiento no es más sorprendente que la capacidad del mismo para llevar a cabo la operación inversa: la síntesis de dicha imagen a partir de sus componentes. A fin de demostrar esta capacidad de síntesis, presentamos a un ojo un enrejado en movimiento hacia la izquierda y, al otro, un enrejado de igual contraste en movimiento hacia la derecha, dejando así al cerebro del observador que realizara la combinación de ambos. La imagen era idéntica al enrejado de contrafase que se habría formado por combinación de ambos enrejados móviles mediante un espejo azogado a medias. Esta síntesis corrobora la idea de que el sistema visual puede ensamblar, como puede separar, imágenes complejas.

La mayoría de los trabajos que hemos expuesto hasta ahora se centraban en la función que desempeñan los mecanismos nerviosos humanos en la detección del movimiento, es decir, en determinar que sea visible o no una "imagen-objeto" dada que no esté fija. Pero aún nos queda un problema todavía más difícil e intrigante: la contribución de los mecanismos mencionados en el aspecto que presentan las "imágenes-objeto" en movimiento, cuando ya se ven con claridad. Los recientes trabajos que hemos realizado en la Northwestern University demuestran que, cuando un observador mira una "imagen-objeto" en movimiento, el que ésta sea visible o no depende de las imágenes percibidas por el observador en un pasado inmediato y de este pasado perceptual también depende el aspecto de la imagen, si ésta llega a hacerse visible. Por ejemplo, hemos comprobado que la adaptación a una lámina

de puntos aleatorios que se mueve uniformemente en una dirección puede alterar, en forma drástica, la dirección aparente de unos puntos móviles, presentados más adelante como estímulo de prueba. El giro que se observa en la dirección percibida – que puede llegar a alcanzar un arco de 10 grados– es máximo cuando la dirección de prueba es similar a la dirección adaptadora; el giro resultante siempre es tal que la dirección percibida se aleja de la dirección adaptadora. Es fácil comprender este giro de la dirección si se supone que el código nervioso que determina la dirección percibida depende de un conjunto de mecanismos específicos para la dirección, en el que cada uno de ellos es sensible a una dirección ligeramente distinta. Los puntos del dibujo adaptador ocasionan un cambio en la distribución de la sensibilidad en el conjunto de los mecanismos, alterando de esta forma la dirección que parece presentar el estímulo de prueba.

Ahora se ha visto ya claramente que, con una ligera estimulación en el umbral de contraste, el sistema visual descompone los estímulos complejos y da una respuesta a cada uno de sus componentes sin que haya interferencia con los demás componentes. No obstante, estos procesos de separación sólo constituyen una verdad parcial. En la vida diaria, la mayoría de los estímulos están muy por encima del umbral de contraste y resulta que el procedimiento que utiliza el sistema visual varía: en lugar de las respuestas independientes, las interacciones son la norma, y tienden a ser inhibitorias.

A fin de poder comprobar como se lleva a cabo dicha inhibición, estudiaremos ahora lo que ocurre en la cacerola de las Molucas, *Limulus polyphemus*. Se sabe mucho acerca de la actividad inhibidora en el ojo compuesto de *Limulus*. Por ejemplo, se sabe que, si uno de los fotorreceptores del ojo compuesto recibe una luz intensa mientras el receptor vecino recibe una luz débil, el receptor más intensamente iluminado reduce la respuesta del receptor vecino. La inhibición es recíproca: cada fotorreceptor puede inhibir a sus vecinos y ser inhibido por ellos. La inhibición mutua neta resultante depende de la intensidad relativa de la iluminación y de la distancia entre uno y otro receptor.

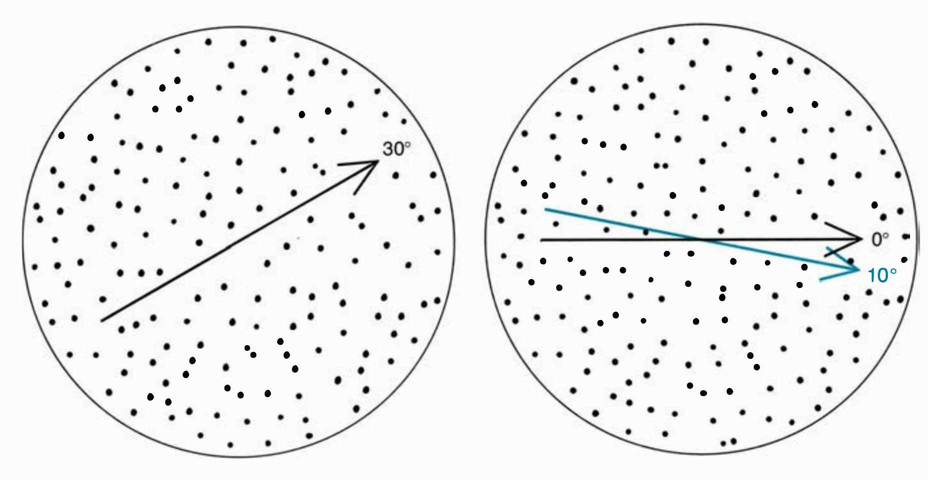
Esta inhibición presenta una característica importante: no siempre se produce. Por ejemplo, hay niveles de iluminación que son adecuados para estimular un fotorreceptor del ojo de *Limulus*, pero no para permitir a ese receptor que inhiba los receptores vecinos. Se debe aumentar la estimulación para que la inhibición se haga patente. Resulta pues que cuando los niveles de estimulación son bajos, cada receptor actúa como una entidad independiente, mientras que, cuando los niveles de estimulación son mayores, se ponen de manifiesto interacciones inhibitorias.

¿Cuál es el papel que podría desempeñar la inhibición en la percepción del movimiento en el hombre? Fred Attneave, de la Universidad de Oregón, ha señalado que la inhibición podría representar un papel en la percepción de distintos dibujos reversibles o multiestables. Como un buen ejemplo de este fenómeno, cita el cubo de Necker, una ilusión visual que parece alternar entre dos orientaciones distintas. También hay dibujos reversibles en la percepción del movimiento, y uno de ellos es el enrejado de contrafase. Ya mencionamos que el enrejado de contrafase parece pulsátil o intermitente cuando se mira durante breves instantes. Sin embargo, cuando se observa durante un período de tiempo algo más largo un enrejado de contrafase de baja frecuencia, su apariencia varía de una forma que recuerda a los cambios que se dan en los dibujos reversibles corrientes. A veces, el enrejado de contrafase parece moverse en una dirección y luego en la otra. El enrejado produce una percepción multiestable y la multiestabilidad, como lo ha demostrado Attneave, es un síntoma de inhibición.

Nosotros también hemos medido de forma más objetiva la inhibición recíproca entre vías sensibles a direcciones opuestas del movimiento. En un experimento, medimos el umbral de contraste para un enrejado que se movía hacia la derecha, después de una adaptación a dos diferentes tipos de estímulo. Uno de ellos era sencillamente un enrejado en movimiento hacia la derecha; el otro era una combinación del mismo enrejado en movimiento hacia la derecha con uno en movimiento hacia la izquierda. Sabíamos que el enrejado adaptador en movimiento hacia la derecha, aislado, disminuiría la visibilidad de un enrejado de prueba que se moviera en la misma dirección. Queríamos comprobar si un componente adicional en movimiento hacia la izquierda reduciría la eficacia del estímulo adaptador en movimiento hacia la derecha. Si al añadir cantidades crecientes del componente en movimiento hacia la izquierda aumentara la inhibición, necesariamente debería quedar



ESTIMULO ESPECIAL **utilizado por los autores en su laboratorio para estudiar la percepción del movimiento en condiciones controladas, y que consiste en un dibujo en movimiento, constituido por puntitos brillantes generados electrónicamente, con la ayuda de una computadora, sobre la pantalla de un tubo de rayos catódicos. Al observador le parecerá una lámina inmensa de puntos aletatorios móviles, vistos detrás de una abertura estacionaria.**



LA ADAPTACION A UNA ZONA DE PUNTOS ALEATORIOS que se mueven uniformemente en una dirección (*izquierda*) **puede alterar de forma drástica la dirección que parece presentar un área de puntos móviles, ofrecida posteriormente como estímulo de prueba** (*derecha*). **El giro de la dirección percibida** (*flecha de color*) **puede alcanzar un arco de 10 grados. El giro resultante es siempre tal que la dirección percibida se va alejando de la dirección adaptadora.**

disminuido el poder adaptador del componente en movimiento hacia la derecha. Resulta que, de hecho, eso es exactamente lo que ocurre.

Los trabajos de fisiología sobre la sensibilidad al movimiento en animales inferiores constituyen un complemento a nuestras observaciones sobre la inhibición de los mecanismos humanos destinados a la percepción del movimiento. En general, la inhibición desempeña una función de "aguzamiento nervioso"; en otras palabras, acentúa las diferencias entre las respuestas de un mecanismo a distintos estímulos. Las reacciones de las neuronas selectivas para la dirección, en la corteza visual del gato, ante direcciones de movimiento que no sean aquellas para las que son selectivas, suelen ser inferiores a la actividad que mantienen estas mismas células cuando no reciben estímulo alguno. Varios investigadores han demostrado hace poco que, en algunos casos, se puede anular esta depresión de la respuesta mediante la administración de la droga bicuculina, una sustancia antagonista del ácido gammaaminobutírico que, según se cree, es un neurotransmisor de tipo inhibidor. En nuestro caso, la inhibición aumenta la diferencia entre la respuesta de una neurona al movimiento en la dirección para la que es selectiva y su respuesta al movimiento en la dirección contraria, mejorando así la selectividad direccional.

Según parece, la inhibición que se da entre unidades específicas para la dirección es débil en el momento de nacer, y su desarrollo requiere un mínimo de experiencia visual. John D. Pettigrew, que actualmente trabaja en el Instituto de Tecnología de California, ha comprobado que dicha inhibición falta totalmente en la respuesta visual al movimiento, cuya dirección no es la específica, tanto en la corteza de un gato recién nacido como en la de un gato adulto privado de visión hasta el momento en que se estimula. Los experimentos de Max Cynader, realizados en la Universidad de Dalhousie, concuerdan con esta observación. Cynader y sus colaboradores criaron gatos en un entorno visual en el que los animales eran sometidos a una sola dirección de movimiento. La gran mayoría de las neuronas corticales selectivas para la dirección, que poseían estos gatos "unidireccionales", eran específicas para el movimiento en la dirección a la cual los gatos habían sido sometidos durante su crecimiento. Esta predisposición referente a la especificidad direccional se contrapone de plano a la distribución relativamente uniforme de la especificidad direccional en los gatos criados en condiciones normales. Aunque aún no se sabe de qué defectos perceptivos puede ir acompañada esta distribución anormal de la especificidad direccional, está claro que las características fisiológicas de los mecanismos destinados a la percepción del movimiento dependen, de una forma crítica, del tipo de mundo visual del cual adquiere experiencia el gato cuando es joven. Se puede especular que la sensibilidad relativamente uniforme que posee el hombre frente a diferentes direcciones de movimiento podría ser una consecuencia de la exposición frecuente, durante la infancia, a movimientos en todas las direcciones.

Hemos mencionado anteriormente que podrían existir vías visuales distintas para el análisis de las características del movimiento y para el de las características de la forma de un mismo estímulo. Hemos demostrado hace poco tiempo que, entre estas distintas vías, sólo las que están implicadas en la percepción del movimiento muestran selectividad direccional. Medimos los umbrales de contraste tanto para la percepción del movimiento como para la percepción de la forma después de una adaptación al movimiento en una dirección; comprobamos asimismo que aumentaban más los umbrales para el movimiento cuando el estímulo de prueba y el estímulo adaptador se movían en la misma dirección que cuando se movían en direcciones opuestas. En cambio, los umbrales para la forma eran iguales fuera cual fuera la dirección del estímulo de prueba. En otras palabras, aunque el estímulo adaptador estuviera ejerciendo un fuerte efecto de selectividad direccional sobre el sistema de percepción del movimiento, no tenía ningún efecto direccional sobre el sistema de percepción de la forma.

Estas observaciones sugieren que el mecanismo nervioso para la detección del movimiento es independiente del mecanismo para la detección de la forma. Este punto de vista está apoyado por otros múltiples experimentos. Ya hemos señalado que las características del movimiento y de la forma de una "imagen-objeto" en movimiento podían presentar umbrales de contraste diferentes. Allan Pantle, de la Universidad de Miami, ha diferenciado ambos mecanismos por un método diferente, mediante la comparación de sus respectivas reacciones ante variaciones de contraste. Se basó en diversos procedimientos para determinar como respondían los mecanismos humanos, sensibles a la dirección, a variaciones del contraste de una "imagen-objeto" en movimiento. Todos los procedimientos coincidieron en el resultado: ambos mecanismos tienen un estrecho margen de respuesta al contraste del estímulo. Su respuesta aumenta linealmente frente al logaritmo del contraste del estímulo en la zona de bajo contraste, pero es prácticamente independiente del contraste del estímulo cuando el contraste supera –por un factor de más de cinco o seis– al nivel umbral. La respuesta de los mecanismos sensibles a la forma no parece saturarse con niveles bajos de contraste.

El examen de los casos de patología cerebral puede dar pruebas adicionales de la separación de estos dos sistemas. Whitman A. Richards, del Instituto de Tecnología de Massachusetts, ha empezado a reunir pruebas de que algunas personas, cuya visión de las formas es normal, tienen dificultades para ver algunos tipos de movimiento. Aún queda por dilucidar si estas dificultades tienen

una base genética, pero la especificidad de esta incapacidad coincide con la idea de una maquinaria especializada para la percepción del movimiento. Además, como consecuencia de algunos tipos de lesiones nerviosas (que pueden haber sido causadas por heridas de bala en la cabeza, por ejemplo), puede darse el caso de que exista una región ciega en el campo visual, llamada escotoma. El examen de estas personas con lesiones neurológicas nos revela que, en las zonas del campo visual en las que no se pueden ver contornos estacionarios, parece conservarse una sensibilidad residual a los estímulos móviles.

A menudo, las personas que padecen un escotoma lo ignoran; al parecer, el sistema visual hace una extrapolación a partir de las áreas que rodean a la región de función deficiente y rellena ésta con lo que ve en las áreas intactas adyacentes. Hace poco, Paul Tynan y uno de nosotros (Sekuler) descubrieron otra forma de extrapolación, totalmente distinta, que se da en observadores normales y que depende de un estímulo móvil. Para poner de manifiesto este efecto que consiste en complementar la imagen, se hace pasar lentamente por la pantalla de un tubo de rayos catódicos un enrejado de frecuencia espacial baja. Luego se pega en el centro de la pantalla, atravesándola de un lado a otro, una cinta de cartulina negra y opaca cuya anchura sea aproximadamente la mitad de la altura de la pantalla. Puesto que la cartulina es opaca, permanece oculto el enrejado subyacente en la región cubierta por ésta. Sin embargo, un observador normal que mire fijamente el centro de la cartulina negra, no verá franjas de enrejado por encima y por debajo de la parte central tapada; en su lugar, ¡verá un enrejado que cubre la pantalla entera! La parte central ilusoria, en cuanto a luminosidad global, se verá bastante más apagada que las partes superior e inferior, pero se verá como real. Estos "fantasmas" móviles parecen moverse a la misma velocidad y en la misma dirección que el dibujo que provoca su formación. En este aspecto, difieren de otras ilusiones de movimiento conocidas. Además, dos franjas muy estrechas de un dibujo móvil, separadas por una gran distancia, pueden producir enrejados fantasmas móviles que parecen reales y se extienden por todo el espacio en blanco que se encuentra entre ellas. Si el enrejado deja de moverse, desaparece la zona fantasma. La ilusión, que también se puede producir con un enrejado radial en movimiento giratorio, requiere para su formación que haya un movimiento de las franjas del enrejado real *(véase la ilustración de esta misma página).*

Floyd Ratliff, de la Universidad Rockefeller, ha propuesto que el sistema visual quizá economice un poco la cantidad de datos que transmite, extrapolando la información procedente de algunos puntos seleccionados del campo visual. Por eso, un dibujo que sea muy repetitivo se ve como si cubriera todo el campo, incluso cuando, en realidad, el dibujo presenta espacios en blanco. Si la información dirigida al cerebro estuviera comprimida y abreviada por economía en la transmisión, el cerebro tendría que fabricar gran parte de nuestra experiencia visual, trabajando con este fin sobre la base de unos datos limitados. Las redundancias del mundo visual son las que, normalmente, permiten que esta fabricación de la imagen pase inadvertida, pero pueden producirse ilusiones cuando se comprimen algunos estímulos poco corrientes para su posterior reconstrucción. Los fantasmas móviles pueden servir para aclarar la naturaleza del proceso de extrapolación.

Aunque el sistema visual normalmente descarta parte de la información que recibe, la falta de determinadas formas de información deteriorará seriamente su función. La teoría matemática del "detector ideal" demuestra que si un observador no sabe qué "imagen-objeto" ha de esperar ni tampoco dónde ni cuándo aparecerá, no podrá detectarla con tanta facilidad como lo hubiera hecho en caso de saberlo. El detector ideal representa la función óptima de detección que puede realizar cualquier sistema sensorial: cuando un sistema sensorial cualquiera no consigue alcanzar el nivel del detector ideal, puede servir para darnos una idea de como funciona este sistema sensorial. Karlene Ball y uno de nosotros (Sekuler), teniendo presente la posibilidad que acabamos de exponer, han comparado la capacidad para detectar unos puntos móviles obscuros en dos tipos de condiciones: cuando la dirección del movimiento se mantiene constante y cuando



EL EFECTO DE CONTORNOS FANTASMAS es una forma de extrapolación visual que se da en observadores normales y que depende de un estímulo en movimiento. Este efecto fue descubierto hace poco en la Northwestern University por uno de los autores (Sekuler) y su colaborador Paul Tynan. Si se mira fijamente el centro de este dibujo radial, mientras está girando sobre un plato de tocadiscos a 33 $^1/_3$ revoluciones por minuto, se observarán unos contornos apagados, si bien distinguibles, que rellenan el espacio vacío que interrumpe el dibujo. Los contornos fantasmas, que van cambiando de fase tal como sus enrejados vecinos reales, aparecerán con mayor claridad si el disco se observa bajo una luz tenue.

varía de una prueba a otra de una forma que resulta imprevisible. La incertidumbre del observador respecto a la dirección del movimiento produce una reducción acentuada de la capacidad de detección. Se ha obtenido resultados análogos cuando lo que es incierto es la velocidad del movimiento de una "imagen-objeto". Mediante el análisis del efecto de la incertidumbre –de varios tipos y cuantías– hemos mejorado nuestros cálculos acerca de las funciones de sensibilización a la dirección y a la velocidad de los mecanismos sensibles al movimiento. Además, los trabajos sobre la incertidumbre respecto al estímulo nos proporcionan un puente de unión entre los datos de laboratorio y la percepción del movimiento en la vida diaria, en la que la incertidumbre respecto al estímulo es lo suficientemente aguda e importante para poder desempeñar un papel limitador destacado.

Los efectos que hemos observado frente a estímulos visuales en movimiento recuerdan los efectos análogos que se dan en otros sentidos. Por tanto, aunque la percepción del movimiento tenga características propias, interesantes y exclusivas, el estudio de esta capacidad particular también es válido por lo que revela sobre la visión y los demás sentidos, más en general. Por ejemplo, hay otras capacidades visuales que parecen utilizar mecanismos coordinados, estando cada uno de ellos proyectado para llevar a cabo una determinada tarea. En el ojo humano coexisten la sensibilidad extrema ante tan sólo unos fotones de luz y un poder de resolución espacial excelente. Esta coexistencia se hace posible, en parte, por la función de dos tipos de fotorreceptores retinianos: los bastones y los conos. Se conoce la existencia de una división del trabajo similar en el tacto, la audición y el gusto.

Parece muy lógico que el sistema visual humano separe la información referente a la forma y al movimiento como también a las distintas direcciones del mismo. En repetidas ocasiones a lo largo de la evolución de nuestra especie, habrá resultado ser mucho más importante responder a la dirección en la que se movía alguna forma mal definida que apreciar los detalles de sus contornos. Cualquier peatón que haya esquivado el tráfico en una gran ciudad, habrá constatado este hecho y debería estar agradecido de que los mecanismos para la percepción del movimiento y para la percepción de la forma puedan coexistir de manera que permite a cada uno realizar lo que tan sólo él puede cumplir con exactitud.

Exoelectrones

Son electrones emitidos por una superficie metálica reciente. Esta aparece con el desgaste o el agrietamiento, asociados a la fatiga metálica. Los exoelectrones dan pie a nuevos estudios de estos procesos

Ernest Rabinowicz

Los investigadores que se disponen a estudiar las grietas, el desgaste u otros defectos de la superficie de un metal, pueden recurrir a una gran variedad de técnicas: microscopía óptica, microscopía electrónica, microscopía electrónica de barrido y autorradiografía, entre otras. Así, un doctorando, al presentar su tesis sobre algún aspecto de la ciencia de superficies, normalmente empieza mostrando una colección de fotografías muy bonitas, y, a renglón seguido, encuentra enormes dificultades para extraer de ellas cualquier dato que pueda ser representado gráficamente y comparado críticamente con otros resultados análogos. Por otra parte, no puede asegurarse que las regiones visibles en las fotografías sean verdaderamente representativas de la totalidad de la superficie. Esta situación poco satisfactoria está siendo superada por una nueva técnica basada en la detección de exoelectrones, que son los electrones emitidos por los átomos de la superficie bajo ciertas condiciones que proporcionan la energía necesaria para provocar su liberación. Resulta que estas condiciones están directamente relacionadas con los cambios en la superficie producidos por el desgaste, las grietas y la fatiga en general. Los exoelectrones pueden contarse fácilmente, de modo que esta técnica proporciona lo que falta en las fotografías: números que puedan ser representados y comparados.

UNA GRIETA DEBIDA A LA FATIGA, en el aluminio, presenta este aspecto en el microscopio óptico. Aunque las imágenes ópticas y otras técnicas normalizadas suelen ser eficaces para revelar grietas y otros defectos superficiales, no indican si son recientes ni cómo avanzan.

El descubrimiento de los exoelectrones estuvo relacionado con el mal funcionamiento de los detectores Geiger, empleados para detectar la radiación, en particular, los electrones producidos en reacciones nucleares o en la desintegración de núcleos radiactivos. Los detectores Geiger son fundamentalmente tubos metálicos, cuyas dimensiones y forma semejan las de un gran cigarro, con una varilla metálica aislada situada en el eje, y una lámina delgada que hace las funciones de ventana en un extremo. Se aplica una diferencia de potencial elevada, generalmente de 1500 voltios, entre el centro de la varilla y la base del cilindro. Si un electrón de alta energía atraviesa la ventana, provoca una descarga eléctrica que puede registrarse y contarse. Ya en los primeros tiempos de utilización de los contadores Geiger, por los años veinte, se observó que un tubo recién fabricado daba un número de cuentas excesiva-

mente elevado durante las primeras horas o días, para pasar a funcionar luego con normalidad. Algo más tarde, se vio que, cuando se introducía en un contador Geiger cualquier superficie fabricada recientemente, crecía también anormalmente el número de cuentas.

El fenómeno fue cuidadosamente investigado, por vez primera, por el físico alemán Johannes Kramer, hacia 1940. Demostró que todas las superficies metálicas, recién preparadas, tenían la propiedad de liberar electrones, y que si estos electrones se emitían dentro de un contador Geiger, provocaban una descarga eléctrica. Kramer encontró que los electrones emitidos tenían una energía de, aproximadamente, 1 electrón-voltio, y que la emisión ocurría solamente durante algunas horas o días después de construida la superficie. De esta manera, se explicó satisfactoriamente el aumento inicial del número de cuentas en los nuevos detectores Geiger.

El anuncio de este descubrimiento sorprendió a la comunidad de físicos. Las condiciones que hacían posible la emisión de electrones de una superficie metálica habían sido establecidas hacía muchos años, como resultados de estudios que ya eran clásicos, y el fenómeno descubierto por Kramer no encajaba. En pocas palabras, se sabía que, para extraer electrones de cualquier superficie, tenía que suministrarse una cantidad específica de energía. Para un metal como el cobre, la extracción de un electrón requiere la aportación de 4,3 electrón-voltios; por ello se dice que el trabajo de extracción del cobre es de 4,3 electrón-voltios. Un método bien conocido para extraer electrones de una superficie es calentar la superficie a una temperatura alta, lo que permite a los electrones adquirir, por excitación térmica, la energía suficiente para superar el trabajo de extracción y abandonar la superficie. Este proceso se conoce como emisión termoiónica. Otro método consiste en iluminar con radiación ultravio-

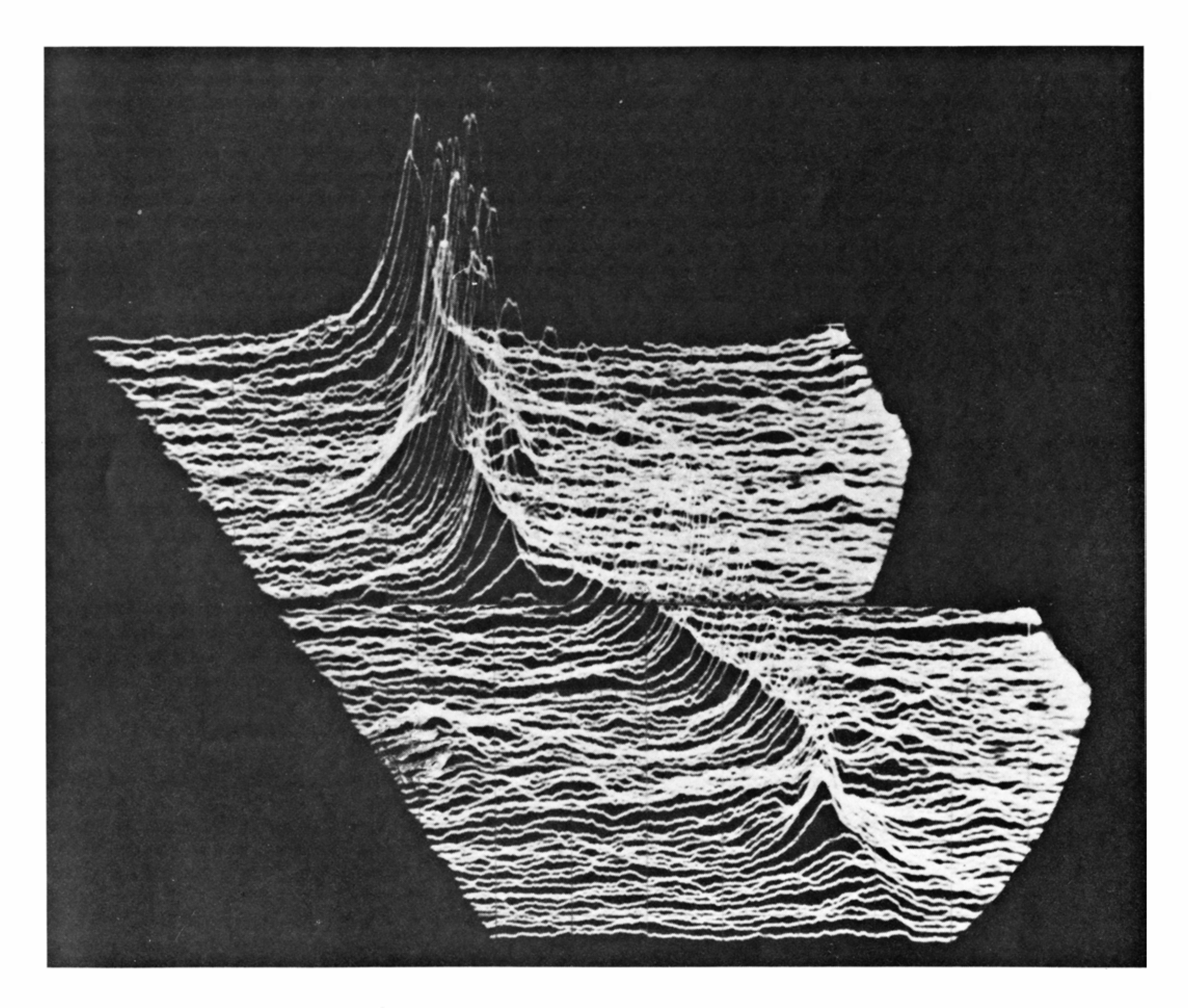
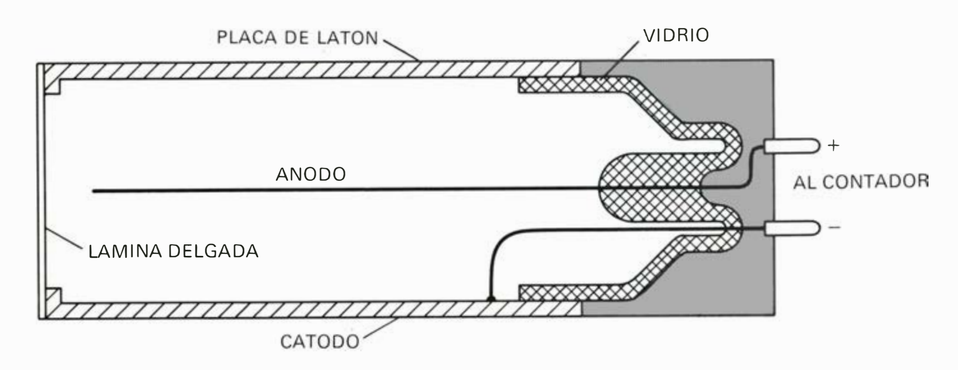


IMAGEN MEDIANTE EXOELECTRONES de la grieta por fatiga mostrada ópticamente en la figura anterior. Esta imagen proporciona una información más gráfica y, potencialmente, más útil. En general, cuanto más reciente es la superficie mayor número de exoelectrones emite. La imagen por exoelectrones de esta figura se obtuvo mediante un barrido de la superficie del aluminio con un haz estrecho de radiación ultravioleta, que estimula la emisión de exoelectrones. La intensidad de la emisión se observa directamente en un tubo de rayos catódicos, en donde el haz de electrones repite el barrido efectuado sobre la superficie metálica por el haz ultravioleta. La desviación vertical en el tubo de rayos catódicos es proporcional a la intensidad de la emisión exoelectrónica. Ambas imágenes fueron obtenidas por C. C. Veerman.



EL COMPORTAMIENTO ANOMALO DE LOS DETECTORES GEIGER condujo al descubrimiento de los exoelectrones. El detector Geiger es, seguramente, el detector de partículas subnucleares mejor conocido; consiste en un cilindro metálico que rodea una varilla delgada de metal, aislados entre sí, de manera que se pueda aplicar una diferencia de potencial de unos 1.500 voltios entre ambos. Cuando un electrón u otro tipo de partícula cargada penetra en el cilindro a través de una lámina delgada que obtura uno de sus extremos, produce una breve descarga eléctrica. Estas descargas pueden ser registradas y contadas. Se observó que los detectores Geiger recién fabricados presentaban, a menudo, un número de descargas anómalas que era especialmente elevado durante las primeras horas o días de funcionamiento.

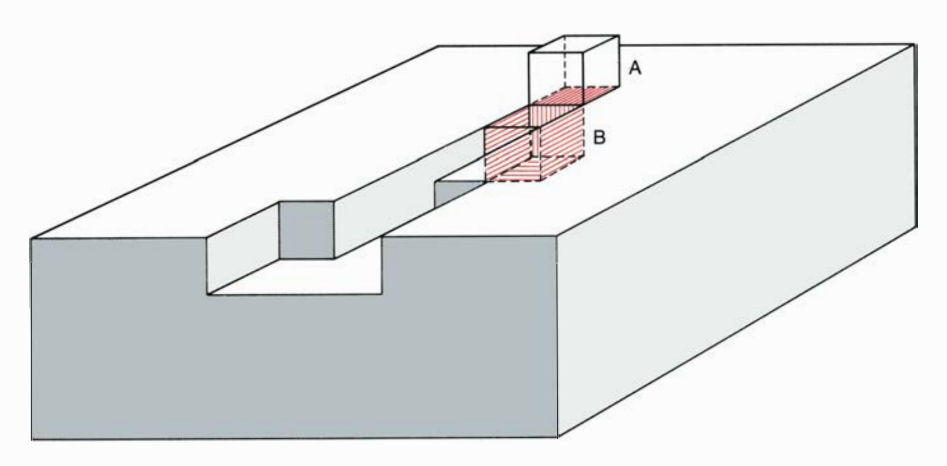
leta la superficie. Si la longitud de onda de la radiación es suficientemente corta, de suerte que la energía de cada fotón supere el trabajo de extracción de un electrón, un fotón puede transferir su energía colisionando con un electrón, y el electrón sale de la superficie. Este proceso se conoce como emisión fotoeléctrica. (Cuando se concedió en 1921, el premio Nobel a Albert Einstein, debióse a su trabajo explicativo del efecto fotoeléctrico, publicado en 1905.)

El descubrimiento de Kramer de la emisión espontánea de electrones resultaba sorprendente, porque describía el carácter espontáneo de un hecho para el que se necesitaba considerable energía. Una buena analogía podría ser una urbanización en una parte del país en la que el agua subterránea estuviera a 4,3 metros por debajo del nivel de superficie. Para obtener agua, cada propietario debería perforar un pozo de 4,3 metros, por lo menos, y bombearla, luego, desde esa profundidad. Llega entonces alguien, abre un pozo nuevo y encuentra que el agua sale espontáneamente del suelo, en chorro, hasta un metro de altura.

¿Qué mecanismo proporciona la energía requerida para superar el trabajo de extracción y causar la emisión de un electrón? Kramer pensó que la energía era de naturaleza térmica. Por ejemplo, si se desgasta una superficie de un metal antiguo, desaparece su capa de óxido y aparece una superficie metálica nueva. La superficie reciente empezará, inmediatamente, a oxidarse de nuevo, liberando una cantidad considerable de energía química en forma de calor. La emisión electrónica se observó también cuando un metal líquido se enfriaba y solidificaba. En este caso, la energía latente de fusión se libera en forma de calor y, cabe pensarlo, parte de la energía térmica liberada pueda dar lugar a la emisión de electrones. Como estos procesos son exotérmicos, es decir, emisores de calor, Kramer llamó a estos electrones exoelectrones. Aunque su explicación del mecanismo de la emisión electrónica no se acepta hoy en día (por ejemplo, se cree que la emisión de exoelectrones en la solidificación está asociada a los cambios de volumen y a la rotura de las capas superficiales que la acompaña), el término exoelectrón ha permanecido. Los exoelectrones se llaman también, algunas veces, electrones Kramer.

El libro de Kramer, publicado en 1950, y en el cual se describen sus descubrimientos, suscitó gran interés entre los investigadores interesados en superficies, quienes intuyeron que los exoelectrones podían serles de gran utilidad en su trabajo. Los físicos hicieron también un esfuerzo ponderable al tratar de explicar cómo y por qué se producían los exoelectrones. A consecuencia de todo ello, los exoelectrones fueron pronto clasificados entre una amplia variedad de fenómenos de superficie a estudiar: fricción, desgaste, corte, molienda de bolas, catálisis, corrosión, fractura, deformación plástica y muchos otros. Además, hay numerosos estudios que utilizan técnicas físicas y químicas para elucidar las propiedades de los exoelectrones y descubrir el mecanismo de su producción. Algunos trabajos de éstos se han publicado, otros no.

Pasemos ahora a examinar los aspectos científicos y de tipo técnico de los exoelectrones. Por lo que respecta a la producción de exoelectrones, no hay duda de que, si una superficie metálica fabricada recientemente desprende un electrón, tiene que haber una fuente de energía. Una superficie fresca tiene dos fuentes de energía evidentes de las que carece otra superficie antigua. Una es la energía química de enlace del metal. Consideremos, de manera extremadamente simplificada, uno de los procesos mediante los cuales puede liberarse esta energía. Cuando se crea una superficie reciente, por un procedimiento cualquiera, pueden quedar algunos átomos en una posición expuesta, de manera que algunos de sus enlaces químicos permanezcan desocupados. Posteriormente, cuando estos átomos formen nuevos enlaces con otros átomos del metal, el

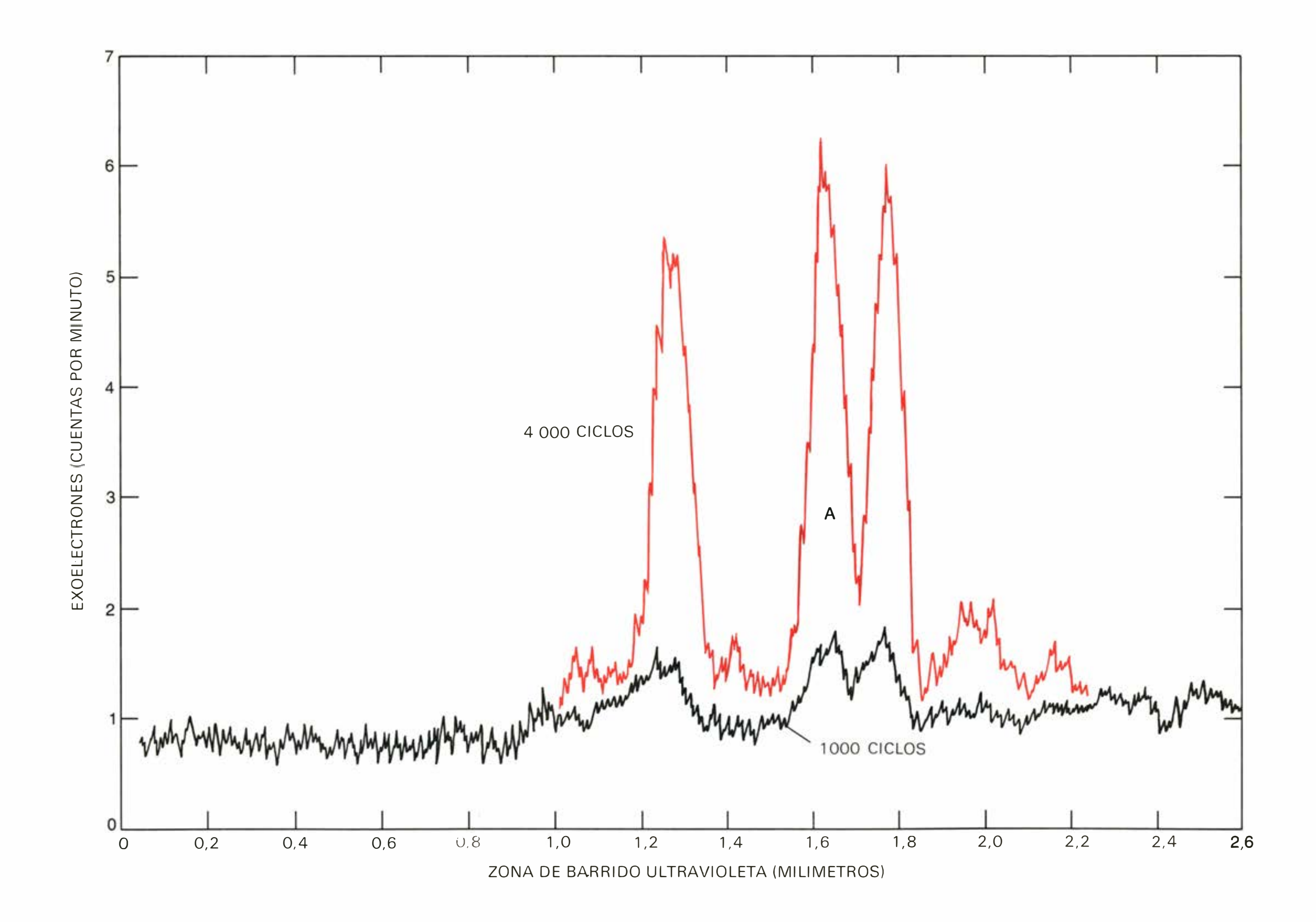


UNA TEORIA DE LA EMISION DE EXOELECTRONES recurre a la energía que se libera cuando un átomo cae desde una posición en la que está ligeramente ligado a la superficie de un metal (*A*) hasta otra posición en la que se encuentra fuertemente ligado (*B*). En este esquema el átomo situado sobre la superficie está ligado en una dirección con el átomo que se encuentra debajo, quedándole enlaces libres en cinco direcciones. Cuando el átomo cae dentro del surco, queda ligado por enlaces en cuatro direcciones, resultando dos únicos enlaces libres. La energía liberada en este desplazamiento equivale a la mitad de la energía necesaria para evaporar el átomo.

número de enlaces desocupados se reducirá y la energía de los nuevos enlaces quedará disponible, bien sea en forma de calor o, tal vez, en forma de exoelectrones. La otra fuente de energía es la que ya he mencionado: la energía de oxidación o de otros procesos químicos análogos.

Podemos estimar la cantidad de energía implicada en cada uno de estos procesos. Como caso extremo del primer proceso, consideremos un átomo de cobre que haya quedado parcialmente sujeto a la superficie, de suerte que se halle ligado solamente por un enlace en una única dirección, en vez de estarlo en seis direcciones como es el caso de los átomos del interior de la red cristalina del metal. Si este átomo cayera en un hueco próximo a la red, podría quedar atrapado por enlaces en cuatro direcciones. En su estado inicial, el átomo de cobre había recorrido cinco sextos del camino a seguir para evaporarse (ya que un átomo de cobre evaporado no está unido con enlace alguno a los otros átomos de cobre); después de caer en el hueco de la red, está como si hubiese recorrido dos sextos de ese camino (ya que hemos supuesto que aún le quedan desocupados los enlaces en dos direcciones), de manera que la energía liberada por el desplazamiento de este átomo resultará ser la diferencia: tres sextos, o un medio, de la energía de evaporación de un átomo de cobre. La energía de evaporación del cobre a la temperatura ambiente es de 2,9 electrón-voltios por átomo, de donde concluimos que la energía neta resultante de este reordenamiento atómico, al ser la mitad de la energía de evaporación, tiene que valer 1,45 electrón-voltios. Consideremos ahora la energía liberada por oxidación. Si un átomo de cobre se oxida a CuO, la energía liberada es de 0,8 electrón-voltios (el cobre también se puede oxidar a Cu_2O con liberación de una cantidad parecida de energía).

Resulta difícil justificar, basándose en estos cálculos, como un exoelectrón puede adquirir los 4,3 electrón-voltios que necesita para superar el trabajo de extracción y escapar de la superficie de un cristal de cobre; y esto, sin considerar la energía adicional, de un electrón-voltio, aproximadamente, correspondiente a la energía cinética típica de un exoelectrón. Entre las explicaciones válidas, una pudiera ser que un átomo de cobre se oxide, adopte una posición en la red con mayor número de enlaces y adquiera una cantidad excepcionalmente elevada de energía térmica, todo ello al mismo tiempo. Otra sería que, en cuanto el cobre empiece a oxidarse, se forme en la superficie una película de óxido muy delgada con un trabajo de extracción reducido, de dos electrón-voltios por ejemplo.



CRECIMIENTO DE GRIETAS en una tira de aluminio, puestas de manifiesto en un estudio por emisión exoelectrónica por William J. Baxter. La probeta fue sometida a esfuerzos repetidos de flexión en el vacío. Periódicamente, se estimuló la emisión de exoelectrones enfocando un haz de radiación ultravioleta. Las gráficas muestran el incremento de la emisión después de 1000 y 4000 ciclos de esfuerzo. La muestra se rompió por la posición en que aparece el pico *A*, después de 140.000 ciclos. Los exoelectrones pueden revelar las grietas en formación.

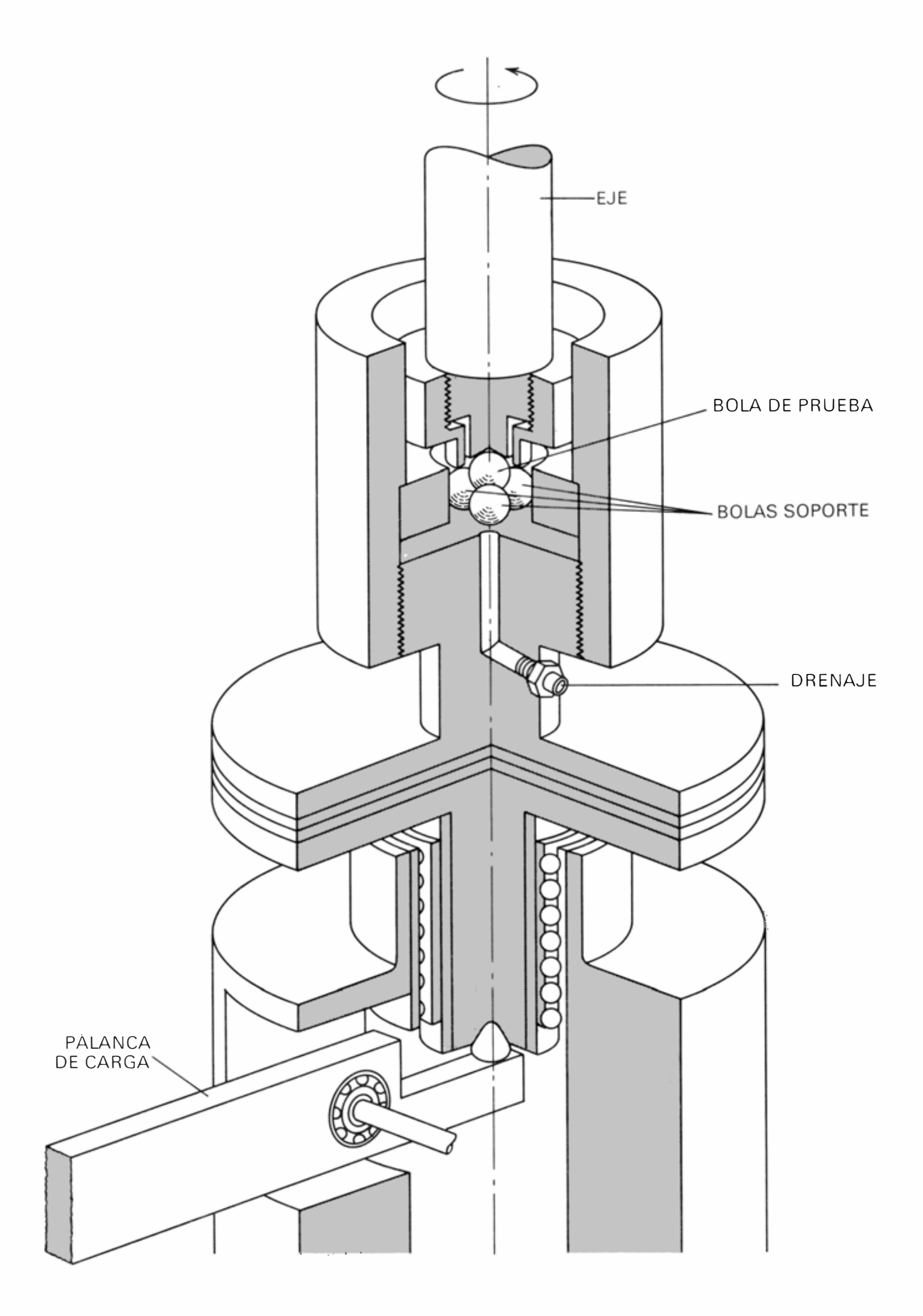
Ello haría más fácil la emisión de un exoelectrón, ya que la energía requerida quedaría reducida, pero, al mismo tiempo, los exoelectrones tendrían que atravesar la capa de óxido antes de alcanzar la superficie. Otro fenómeno posible que debe considerarse es el de la adsorción, en la superficie del sólido, de moléculas gaseosas del medio que le rodea, Dicha adsorción puede proporcionar energía, pero en una cantidad bastante limitada.

Después de haber considerado estas posibilidades, da la impresión de que la emisión de un exoelectrón depende de una combinación excepcional de circunstancias favorables. De hecho, los cálculos basados en la proporción de exoelectrones emitidos confirman la hipótesis de que la emisión de un exoelectrón es un proceso poco probable. Si se desgasta un centímetro cuadrado de la superficie de un metal, la superficie que aparece contiene unos 10^{15} átomos metálicos. Si la emisión de un exoelectrón fuese el comportamiento normal de los átomos de la superficie, podríamos suponer que cada átomo, en la superficie, emite, por lo menos, un exoelectrón; de manera que, al cabo de unos cuantos días, podría haberse emitido exoelectrones por un total de unos 10^{15} por cada centímetro cuadrado de superficie reciente. El número total de exoelectrones es muy inferior, puede alcanzar poco más de 10^{8} por centímetro cuadrado, incluso en circunstancias favorables. Por consiguiente, sólo un sitio en la red superficial, por cada 10 millones de ellos, se encuentra con el conjunto especial de condiciones requeridas para la emisión de un exoelectrón.

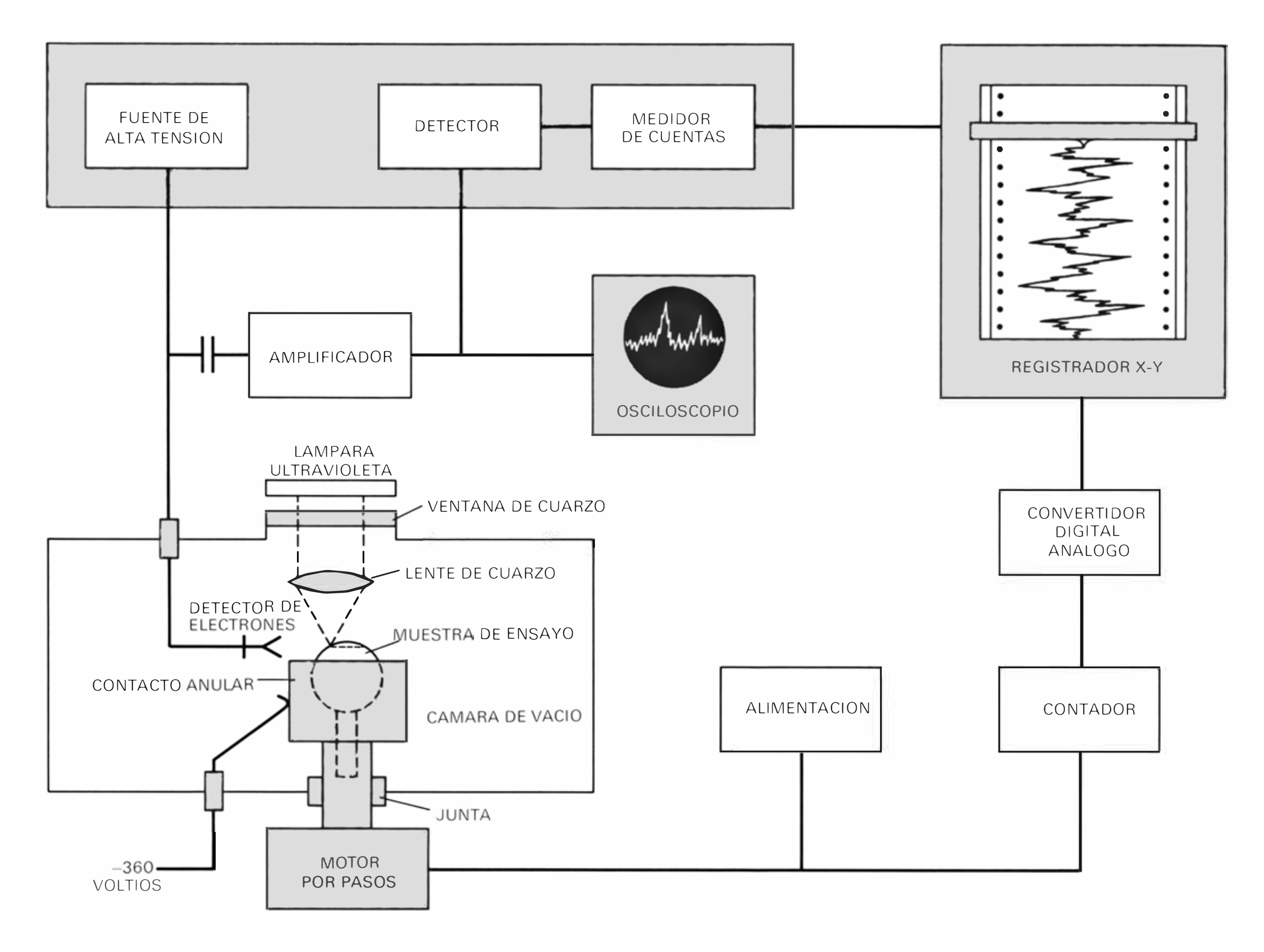
Hasta aquí cada experto en el fenómeno de los exoelectrones parece tener su propia teoría sobre la combinación exacta de condiciones que se necesitan para la emisión exoelectrónica. Constituye un problema el que cada investigador realizara sus experimentos bajo unas condiciones distintas, y, naturalmente, los resultados a que llega no se parecen a los conseguidos por otros investigadores. Resulta verosímil que distintas combinaciones de energía mecánica disponible, energía química, energía de adsorción y energía térmica hayan de intervenir en distintas circunstancias para compensar el trabajo de extracción, que cambia a su vez como resultado de las reacciones químicas y de la contaminación de la superficie.

Los investigadores de exoelectrones hallaron muy pronto que la emisión de exoelectrones podría proporcionar información sobre muchos de los efectos relacionados con la superficie. Por ejemplo, cuando un material áspero se pone en un molino de bolas (esencialmente un tambor giratorio lleno de bolas de acero) para reducirlo a polvo, al principio hay una reducción continua en el tamaño de las partículas, pero seguidamente la velocidad de reducción decrece, hasta llegar a anularse. El fenómeno está completamente reflejado en la tasa de emisión electrónica de las partículas, que es bastante grande en los primeros estadios del proceso, pero va decreciendo continuamente.

Otro ejemplo considerado era qué ocurre cuando dos metales se deslizan uno sobre el otro bajo distintas cargas: emiten más exoelectrones para cargas elevadas que para cargas pequeñas. Esto apenas si sorprendió, ya que se podría esperar que un área superficial mayor sería perturbada más por cargas grandes que por las pequeñas. Aunque la correlación entre emisión exoelectrónica y carga resultaba interesante, nadie



LA FATIGA DE LAS BOLAS DE LOS RODAMIENTOS se estudia en un aparato en el que una bola de prueba, asentada sobre tres bolas soporte, gira a 3.560 r.p.m. Las bolas se lubrican constantemente con lubricante nuevo. El contacto sobre la bola superior queda confinado a una estrecha pista circular. El esfuerzo sobre la bola de prueba se controla mediante el peso colocado en la palanca de carga. La prueba se detiene automáticamente cuando un micrófono detecta una vibración, señal de un defecto grave en la superficie de una bola.



LA EMISION EXOELECTRONICA de las bolas de los rodamientos se midió y registró con este aparato. Después de retiradas del aparato esquematizado en la página anterior, las bolas se limpiaron y se montaron en una cámara de vacío, de modo que un haz estrecho de radiación ultravioleta incidiera en la pista sometida al ensayo de fatiga; así, al girar la bola lentamente, se recorría toda la zona de barrido. El haz estimula la emisión de exoelectrones. El número de exoelectrones emitidos se observa en el osciloscopio y se representa en un registrador *x-y*.

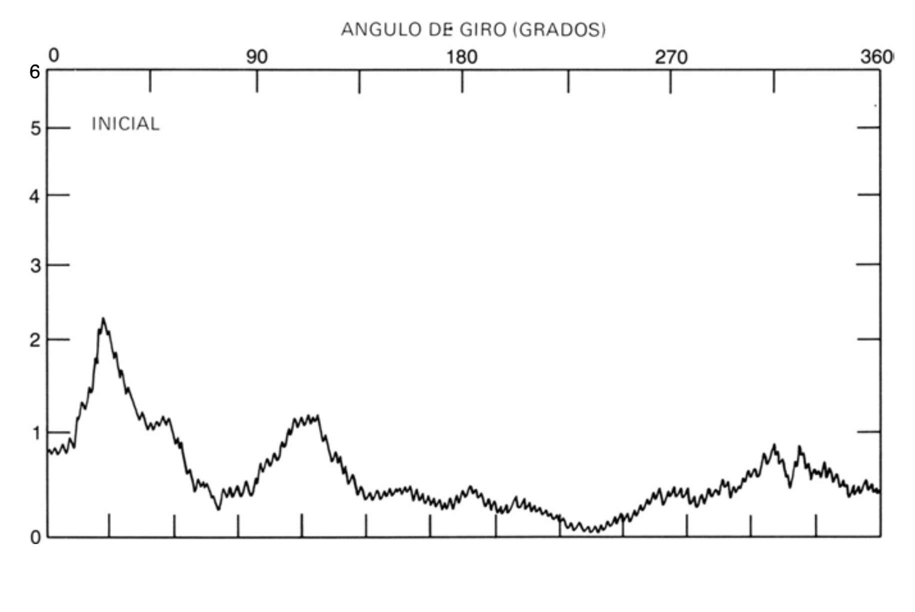
pudo ver ninguna manera de interpretar la emisión para adquirir un conocimiento significativamente nuevo del fenómeno del rozamiento.

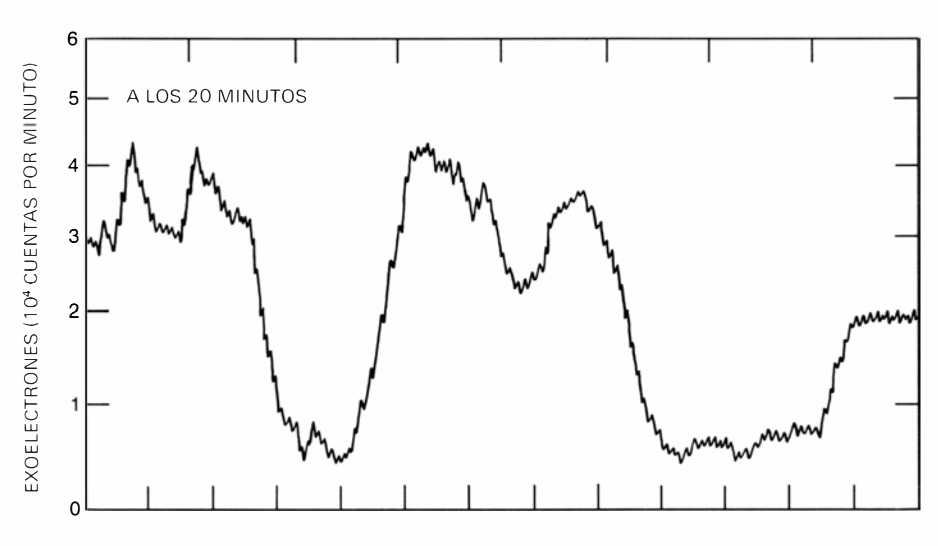
Básicamente, el problema era el siguiente: si los exoelectrones tenían que utilizarse como medio para conocer otros fenómenos, se imponía saber con precisión bajo qué condiciones se producían. Sin embargo, según se vio, aquí había radicado la mayor dificultad. El interés por los exoelectrones alcanzó su grado máximo en 1957, año de una conferencia internacional sobre este tema; a partir de entonces, quizás por no profundizarse en la explicación del mecanismo de su emisión, declinó hasta un punto bajo a mediados de la última década. A finales de la misma, no obstante, el interés volvió a surgir con el descubrimiento de nuevas formas potenciales de utilizar los exoelectrones, algunas de las cuales pasamos a describir.

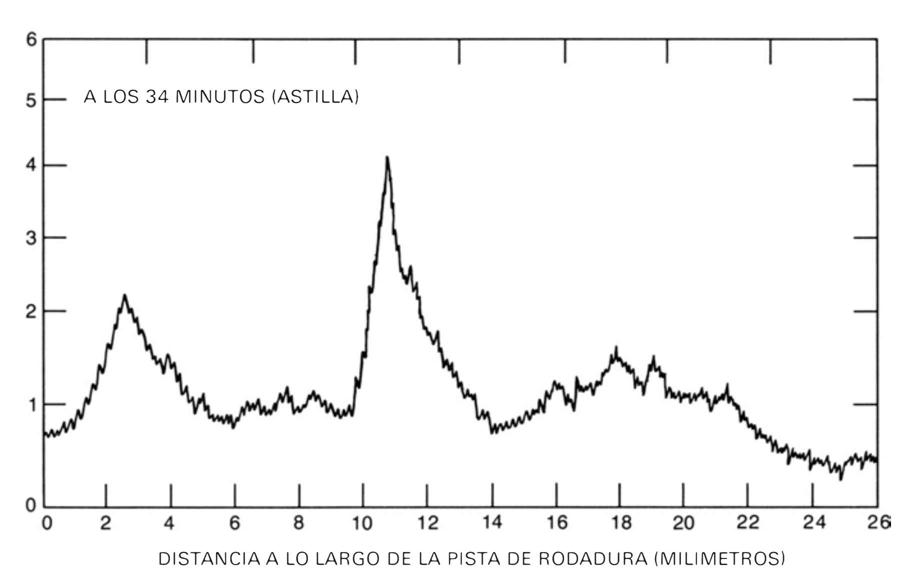
Uno de los nuevos métodos investiga la emisión de exoelectrones como un índice del daño por radiación o exposición a la radiación. Se halló que si ciertos materiales inorgánicos, tales como sulfato cálcico, fluoruro de litio y óxido de berilio, se exponen a radiación nuclear u otros tipos de radiación de alta energía, su estructura cristalina sufre un daño que es proporcional al tiempo de exposición. Este daño puede invertirse recociendo el material a una temperatura determinada, generalmente unos 1000 grados Celsius. Al propio tiempo, se emiten exoelectrones en función, más o menos directa de la radiación absorbida por la muestra. El fenómeno está en fase de estudio dadas sus posibles aplicaciones a un dispositivo recuperable, útil para controlar la exposición a la radiación que sufren los trabajadores de centrales nucleares y de otros lugares similares. El método promete ciertas ventajas sobre otras técnicas de control, las cuales miden la pérdida de carga electrostática o la impresión de una película fotográfica.

Otra aplicación reciente de los exoelectrones consiste en estudiar el proceso de crecimiento de grietas en sólidos, en particular durante la fatiga. La fatiga es una forma de fallo o rotura de sólidos metálicos cuando se les ha sometido a esfuerzos mecánicos cíclicamente variables. Supongamos que una regla metálica se estira y comprime repetidamente. Si los esfuerzos son menores que la fuerza elástica del metal, sobrevivirá a las primeras aplicaciones de tensión sin sufrir daño. Pero si continúan repitiéndose las aplicaciones de tensión, la regla puede romperse como resultado de la fatiga.

Los primeros estudios mostraron que, en estos casos, se forman pequeñas grietas en el metal, que van creciendo gra-







LA HISTORIA DE LA ROTURA DE UNA BOLA DE RODAMIENTO puede reconstruirse observando la emisión de exoelectrones y su localización sobre la bola, durante el ensayo de fatiga. Estos registros se obtuvieron con los aparatos ilustrados en la página precedente. La gráfica superior muestra la emisión de exoelectrones antes de empezar el ensayo. La segunda gráfica muestra la emisión a lo largo de la pista de rodadura, a los 20 minutos de ensayo. La bola se rompió a los 34 minutos, saltando una astilla. El pico agudo de la última gráfica coincide con el lugar dañado.

dualmente durante los ciclos sucesivos de tensión mecánica, hasta que la muestra se rompe en dos. En los primeros estadios de fatiga, además de que las grietas son pequeñas y difíciles de detectar, no existe ninguna posibilidad de predecir cuáles van a crecer y cuáles quedarán estacionarias. Por este motivo, el estudio del crecimiento de grietas fue, durante mucho tiempo, un problema arduo. Sin embargo, por tratarse de una grieta en crecimiento, una superficie reciente habrá de emitir exoelectrones, y así es. De ahí que los exoelectrones se estén utilizando para investigar la formación y el crecimiento de las grietas.

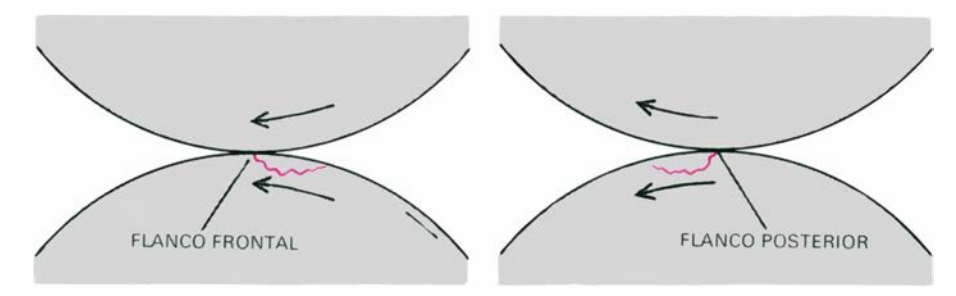
En tales estudios sería interesante poder localizar con precisión dónde se emiten los exoelectrones. Hay un método puramente mecánico de localizarlos. Consiste en desplazar lentamente una máscara con un pequeño agujero a través de la superficie de la muestra, efectuando un barrido de la superficie de tal modo que permita conocer, en cada instante, la fuente de exoelectrones que alcanza el detector. Otra técnica, mucho más depurada, procede de los primeros trabajos con exoelectrones. Entre las técnicas de diagnóstico utilizadas por los primeros investigadores para estudiar el proceso exoelectrónico, una consistía en aplicar distintos tipos de energía (calor, luz, vibraciones mecánicas) a las superficies que desprenden exoelectrones y estudiar las consecuencias. En general, cuando se aplica una energía externa, el número de exoelectrones se incrementa. Los electrones así desprendidos se llaman exoelectrones estimulados. El estimulante más eficaz es la radiación ultravioleta. La frecuencia de la radiación debe ser suficientemente baja para que la energía de cada fotón sea menor que el trabajo de extracción; de lo contrario, habría también emisión fotoeléctrica además de emisión exoelectrónica, perfectamente establecida.

Con la radiación ultravioleta el número de exoelectrones emitidos por una superficie puede incrementarse en un factor de más de 10.000, respecto a la emisión normal. Esto significa que, si una pequeña región de una superficie se ilumina mientras se mide la emisión exoelectrónica de la totalidad, casi toda la emisión proviene del área iluminada. En consecuencia, podemos barrer con un haz de radiación ultravioleta toda la superficie y medir simplemente la emisión de electrones en función del tiempo. El número de exoelectrones contados en los distintos tiempos indica la emisión procedente de las diferentes zonas de la superficie sólida.

Entre las primeras demostraciones de la información que puede obtenerse mediante esta técnica se encuentra la presentada por William J. Baxter, del Research Laboratories of the General Motors Corporation. Colocó una tira de aluminio en una cámara de vacío que contenía un detector de electrones; la sometió a esfuerzos cíclicos, flexionándola. Al propio tiempo, iba barriendo la superficie con un haz de radiación ultravioleta enfocado puntualmente. Antes de aplicar los esfuerzos mecánicos, no había emisión de exoelectrones por parte de la superficie. Después de 1000 ciclos, había tres lugares que presentaban una emisión por encima de la media. Después de 4000 ciclos, los sitios activos se habían hecho más prominentes. Uno de ellos dio lugar a la rotura después de 140.000 ciclos de esfuerzos. Está claro que los sitios activos indicaban los lugares donde se formaban, crecían y se extendían las grietas. El resultado de Baxter era muy interesante, ya que, hasta entonces, no se disponía de ningún método para identificar grietas finas en superficies metálicas y seguir, consiguientemente, su evolución hasta la rotura.

El trabajo de Baxter nos animó, en el Instituto de Tecnología de Massachusetts, a intentar aplicar las mismas técnicas al problema del crecimiento de grietas en rodamientos a bolas, un proceso llamado generalmente desgaste por fatiga superficial. El desgaste por fatiga superficial es la causa principal del fallo de los rodamientos a bolas o cilindros, y, al propio tiempo, causa importante de la fatiga de los engranajes y de las ruedas que circulan sobre raíles. En los rodamientos a bolas, la fatiga superficial se desarrolla cuando se forma una grieta en una de las bolas y crece lentamente hasta que llega a desprenderse una astilla de tamaño apreciable. Una vez ha ocurrido esto, el rodamiento se destruye rápidamente.

El desgaste por fatiga superficial es un fenómeno incómodo de analizar, ya que hay una gran dispersión estadística en el tiempo necesario para que el fallo se produzca. Si se investigan 100 rodamientos a bolas hasta la rotura, los tiempos medios entre el primero y el último fallo pueden diferir en un factor de 100 o más. El problema resulta más difícil todavía, ya que, a escala macroscópica, no existen mediciones útiles que puedan hacerse mientras se espera que la bola se rompa. A escala microscópica, podemos examinar la formación de grietas en las superficies, que son pequeñas y difíciles de localizar. A gran aumento, algunas grietas pequeñas se hacen visibles, pero no existe un método adecuado que nos permita averiguar cuál de ellas está creciendo rápidamente y puede originar una rotura. Por todo ello, nos pareció que los exoelectrones podrían proporcionar una vía para descubrir las grietas significativas y observarlas durante su crecimiento.



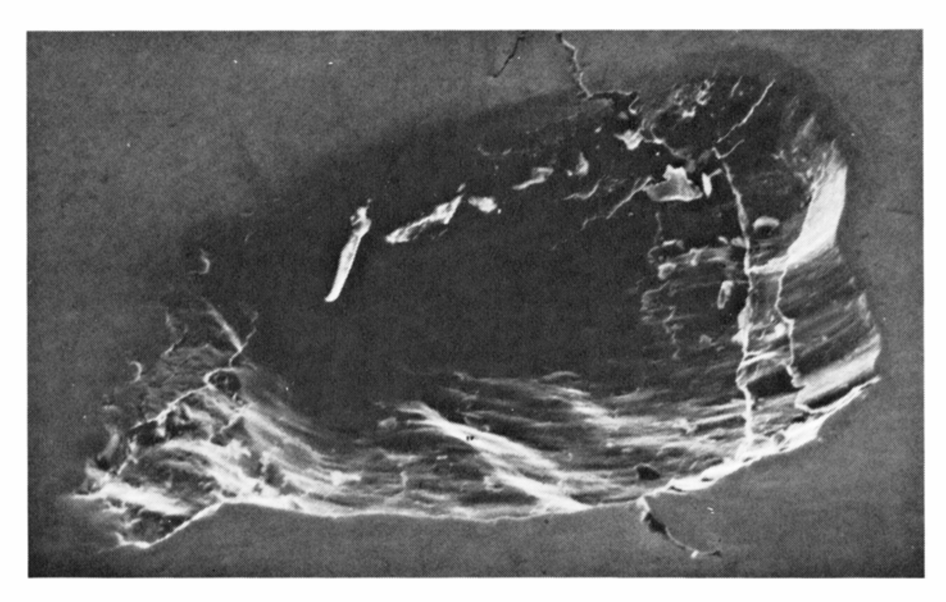
UNA GRIETA EN LA SUPERFICIE de una bola de rodamiento puede progresar hasta desconcharse una astilla. El proceso puede evolucionar a partir del borde de ataque de la futura astilla *(izquierda)*, o bien en dirección opuesta a partir del borde de arrastre *(derecha)*. Antes de las técnicas de observación basadas en la emisión exoelectrónica, sólo existían métodos indirectos para establecer la dirección del crecimiento progresivo de una grieta.

Los problemas experimentales eran graves. Puesto que los electrones viajan en el aire a lo largo de distancias muy cortas antes de ser frenados, resulta virtualmente necesario experimentar con ellos en un sistema de vacío. Por otra parte, para ser significativas, las pruebas con rodamientos a bolas tienen que realizarse en las condiciones en que los rodamientos trabajan normalmente: en aire con superficies lubricadas. En consecuencia, para medir la emisión exoelectrónica de rodamientos a bolas se necesitarían pruebas de funcionamiento en baño de aceite, seguidas de limpieza del aceite con un disolvente, evaporación de este último, colocación de las bolas en la cámara de vacío y realización del vacío en la cámara; luego, hay que esperar que todo esto no interfiera en la emisión de exoelectrones. Cuando el experimento llegó a oídos de mis colegas, algunos pensaron que era altamente improbable que, después de todas estas manipulaciones, pudiera detectarse algún exoelectrón, pero decidimos llevar adelante el proyecto. La U.S. Army Research Office se ofreció para sufragar los gastos de investigación.

Se usaron dos aparatos, una cámara de ensayos de fatigas, en la que los rodamientos a bolas se sometían a tensiones mecánicas, y una cámara de vacío, en donde se detectaba a los electrones. La primera era del tipo de cuatro bolas de Barwell; en ella, una bola de prueba, sostenida por un eje que gira a 3650 revoluciones por minuto, cabalga sobre tres bolas que están obligadas, por un anillo y un disco, a moverse sobre una pista circular. Todas las bolas miden 12,7 mm de diámetro. (Se ha comprobado que los experimentos llevados a cabo en este aparato, con cuatro bolas en rodamiento a la vez, dan resultados que concuerdan con un rodamiento a bolas clásico, el cual consta de un anillo interno, un anillo externo y un gran número de bolas entre ambos.) La sección sujeta a investigación se va lubricando constantemente. Un micrófono aplicado a la cámara mide la vibración y desconecta rápidamente la corriente cuando alguna bola empieza a astillarse. La geometría es tal que el contacto queda restringido a una pista relativamente estrecha de la bola superior, de modo que el daño y la rotura suelen presentarse en esta región. Por ello, sólo debemos examinar, en cuanto a emisión exoelectrónica, a la bola superior.

El aparato de vacío detector del exoelectrones utiliza radiación ultravioleta para estimular la emisión. Los rayos ultravioletas pasan a través de una ventana de cuarzo, y se localizan sobre la muestra mediante una lente de cuarzo. Para los resultados presentados aquí, la medida era de 2 milímetros por 0,27 milímetros. Los electrones emitidos se detectan mediante un multiplicador de electrones. Se suman los pulsos de salida y se representan en el eje y de un registro x-y.

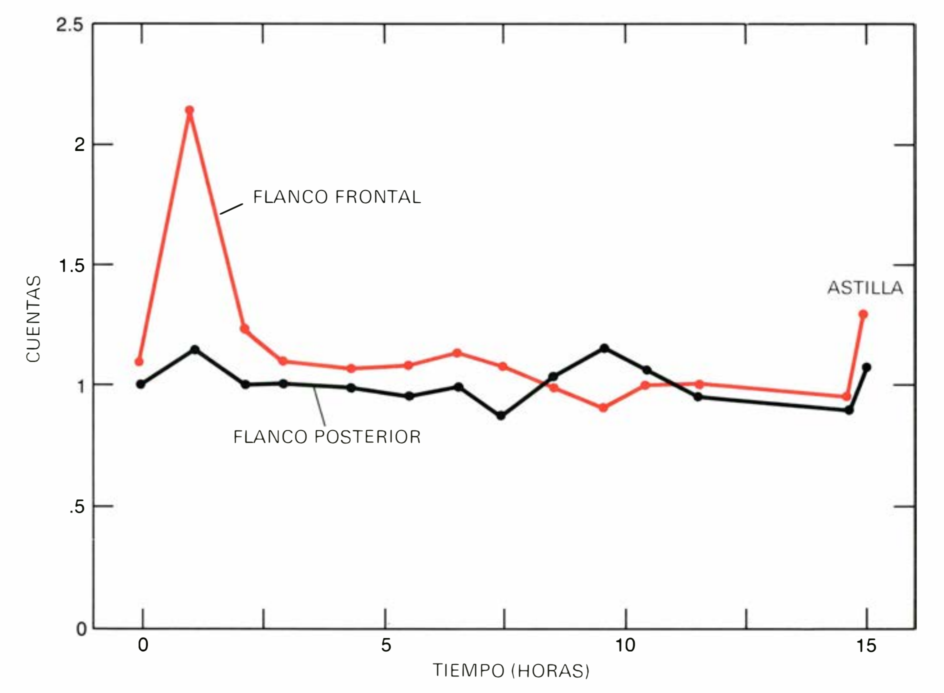
Las mediciones iniciales se hicieron con exoelectrones emitidos por rodamientos a bolas lubricadas, previamente rayadas con una punta de diamante y limpiadas con desengrasante Freon. El detector de exoelectrones trabajó tal como esperábamos, dando máximos acusados cuando el haz ultravioleta iluminaba la región rayada de la bola. Hubo posteriormente una serie de pruebas de desgaste. En una de ellas, las bolas de acero giraban en el aparato de



ASTILLA TIPICA debida a la fatiga en una bola de rodamiento, vista en un microscopio electrónico de barrido. Tiene 1,8 mm de longitud, y el borde de ataque se halla a la izquierda de la fotografía.

Barwell sometidas a un nivel de tensiones tal que la vida media anticipada era de unos 40 minutos. Cada 10 minutos, aproximadamente, se sacaba la bola superior, se examinaba su emisión de exoelectrones y se devolvía al aparato para que continuase sometida a tensiones. La bola de prueba se astilló a los 34 minutos de experimentación. Cuando la zona desgastada de la bola se examinó a los 20 minutos, presentó tres grandes picos de emisión de exoelectrones. El lugar de uno de ellos correspondía al sitio que acabó por astillarse.

Además de permitir localizar la posición de una posible rotura por fatiga superficial, los exoelectrones ponen en claro el mecanismo que sigue el proceso de fallo. Por ejemplo, si un rodamiento a bolas se somete a tensiones en el aparato de pruebas, o durante su utilización, de manera que la bola siempre ruede sobre la superficie adyacente en la misma dirección, una grieta que se desarrolle en su superficie crece típicamente hacia dentro en ángulo, y puede llegar a desprender una astilla. La cuestión que surge es si la ranura crece en la dirección del rodamiento o en la dirección contraria. Problema que puede estudiarse si se observa como varía, con el tiempo, la emisión exoelectrónica cerca de la grieta que crece hasta formar una astilla.



GRAFICAS DE LA EMISION DE EXOELECTRONES obtenidas con la bola que acabó astillándose, mostrada en la figura anterior. Estas gráficas permitieron establecer que la grieta inicial aparecía en el lugar que después se convertía en el borde de ataque de la astilla. La grieta aparece, en un principio, en el borde de ataque, progresa por debajo de la superficie, donde no puede detectarse, y emerge, finalmente, por el borde de arrastre.

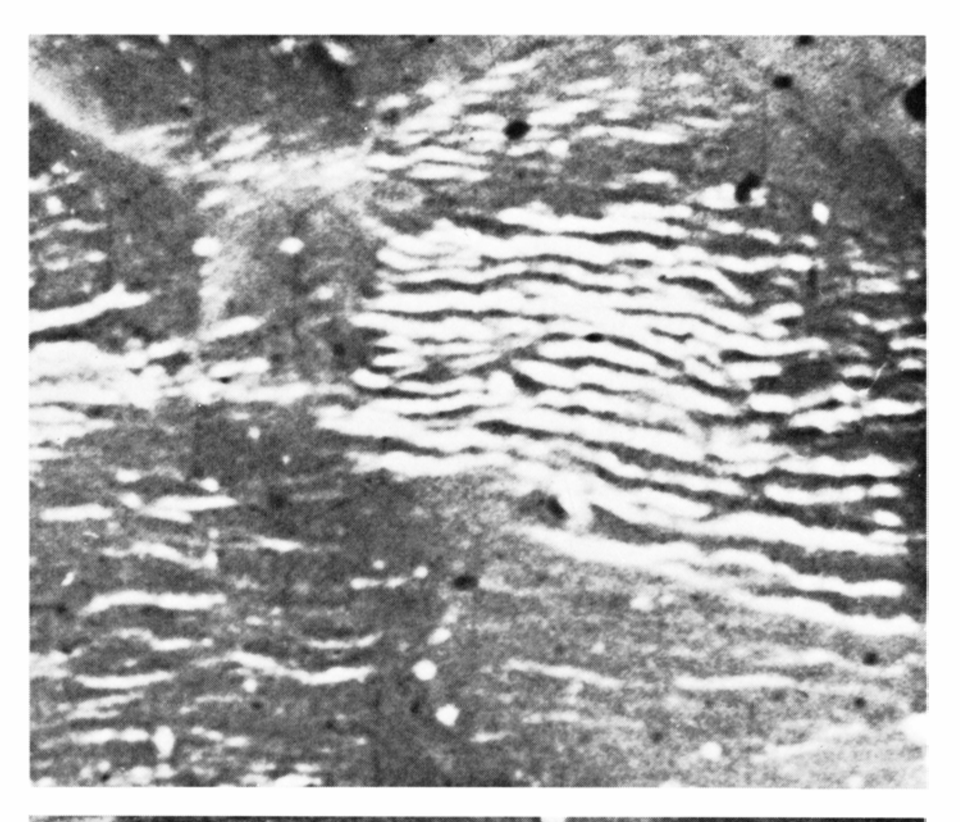
En nuestros estudios hemos conseguido mostrar que el flanco frontal de una posible astilla muestra un pico prominente al principio de la prueba, pero decrece gradualmente, tal vez porque la grieta se extiende por debajo de la superficie y paralelamente a ella. El canto de arrastre muestra raramente una emisión exoelectrónica superior a la de fondo antes de que ocurra la formación de una astilla. Por tanto, la grieta se desarrolla a partir del borde de ataque hacia el borde de arrastre, lo que confirma estudios anteriores sobre este aspecto que permanecían más obscuros.

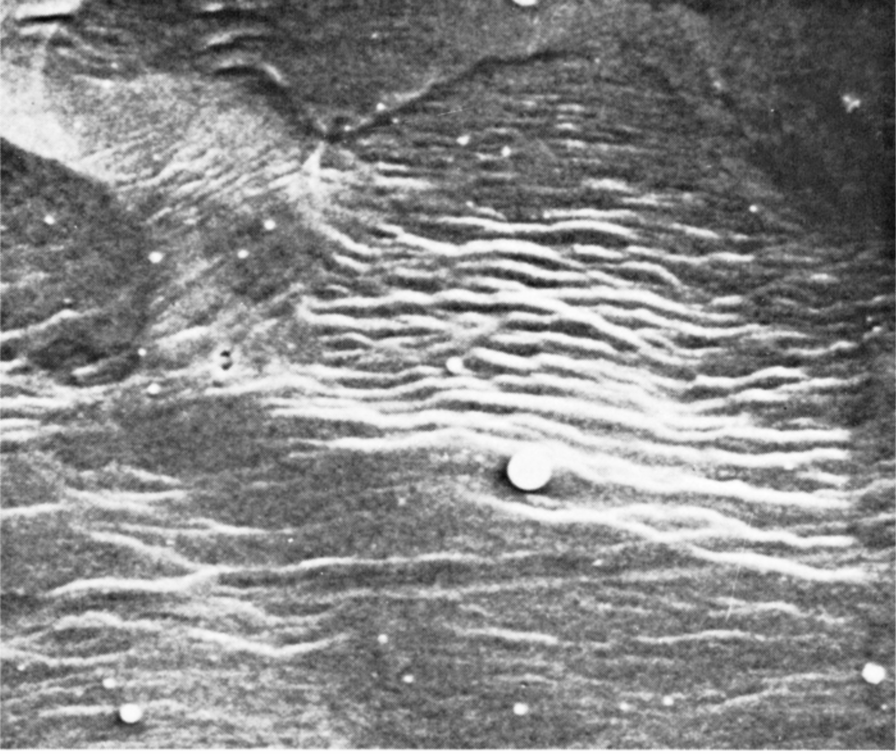
En nuestro trabajo con rodamientos a bolas, el haz ultravioleta puede incidir sobre una sola pista, ya que la zona sometida a tensiones está perfectamente localizada en el aparato de Barwell. No obstante, cabe suponer una técnica en la que la superficie de una muestra queda barrida con un haz ultravioleta, dibujando una retícula de exploración, mientras la emisión de exoelectrones se representa sincrónicamente como desviación vertical de un barrido de electrones sobre un oscilógrafo de rayos catódicos. La primera demostración de esta técnica la realizó, en 1966, por C. C. Veerman, de Delft University of Technology, en Holanda. En las fotografías electrónicas de una grieta por fatiga en una muestra de aluminio, sacadas por él, la grieta se parece a una cordillera de montañas *(véase la ilustración de la página 57)*.

En otro método de barrido similar, la emisión exoelectrónica modula un haz de electrones, análogamente a una imagen de televisión. En tales imágenes, las áreas iluminadas corresponden a las regiones que emiten exoelectrones.

Resulta arriesgado predecir qué pueden aportar los futuros trabajos de emisión exoelectrónica. Dentro de los próximos diez años tal vez se consiga, con técnicas especiales para detectar exoelectrones en aire, examinar sistemáticamente toda la estructura del ala de un avión, investigando el crecimiento de grietas por fatiga. Puesto que las velocidades de los trenes de viajeros crecen cada vez más, sería importante desarrollar métodos rápidos y directos para de-

tectar defectos en las ruedas y en los raíles; quizás las técnicas de exoelectrones ayuden en este propósito. Cabe pensar también que los exoelectrones puedan aplicarse a la predicción de terremotos, al indicar qué grietas de la superficie de la tierra o de las estructuras rocosas del subsuelo se hallan en crecimiento activo. Pero sospecho que, conforme vayamos aprendiendo más de las condiciones exactas en que se realiza la emisión de exoelectrones, iremos encontrando aplicaciones que, hoy en día, no podemos imaginar.





COMPARACION DE DOS TECNICAS para visualizar las grietas en una superficie de acero, sometida a una deformación plástica. La imagen superior es una microfotografía en la cual el brillo es proporcional a la emisión de exoelectrones. La imagen inferior muestra la misma región vista en un microscopio electrónico clásico; las grietas superficiales aparecen como líneas brillantes. Las microfotografías se deben, respectivamente, a Baxter y S. R. Rouse.

Formación de cráteres en el sistema solar

La era de la exploración espacial ha puesto de manifiesto que todos los cuerpos del sistema solar interior poseen cráteres semejantes a los de la Luna. La formación de éstos posibilita el conocimiento de la historia del sistema

William K. Hartmann

Todo planeta y satélite del sistema solar interno está marcado con cráteres. Los cráteres de la Luna fueron descubiertos por Galileo cuando por primera vez apuntó su telescopio hacia el firmamento en 1609. Los cráteres de la Tierra se identificaron en su mayor parte por los geólogos hace pocas decenas de años. Los cráteres de Marte fueron descubiertos por la nave espacial *Mariner 4* en 1965. Los cráteres de las dos lunas de Marte, Phobos y Deimos, fueron fotografiados por el *Mariner 9* en 1971. Los cráteres de la superficie desconocida de Venus, perpetuamente velada por nubes, fueron revelados gracias a levantamientos topográficos realizados por radar en 1972. Finalmente, los cráteres abundan en las fotografías de Mercurio enviadas por el *Mariner 10* en 1974. Ningún planeta o satélite del sistema solar externo se ha fotografiado hasta ahora con poder separador suficiente para revelar si existen o no cráteres en él.

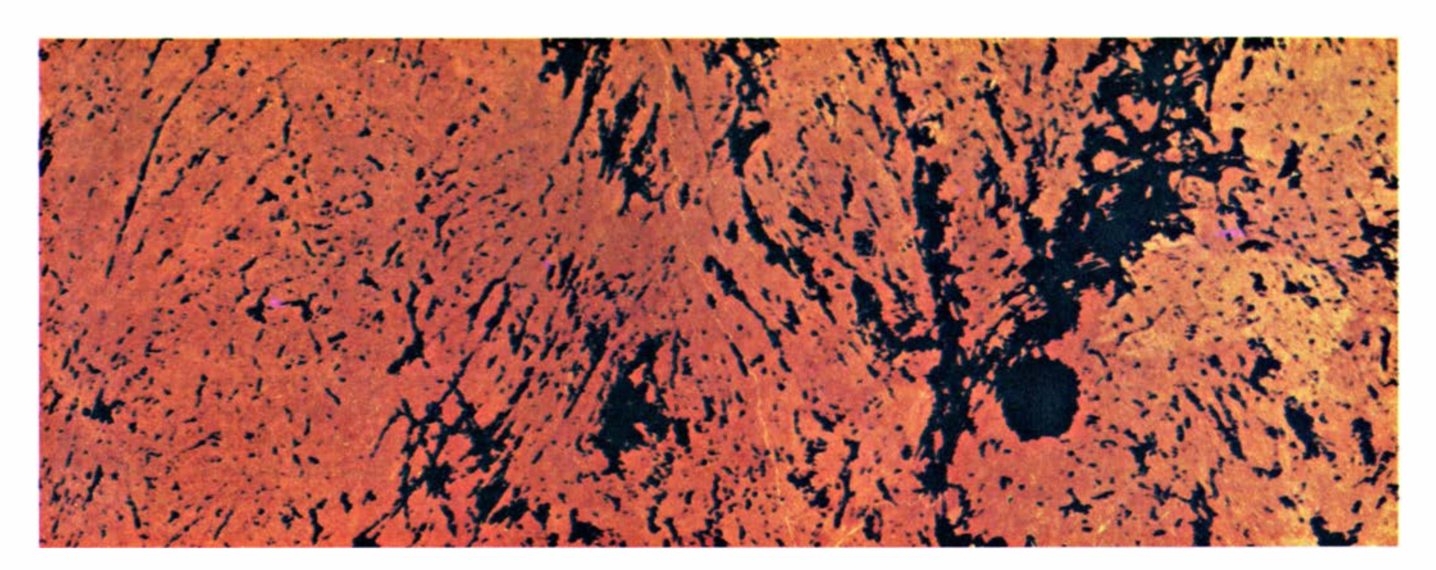
Para quienes estudian el sistema solar, la ubicuidad de cráteres en la superficie de los planetas y satélites interiores es una suerte. La forma y el número de los cráteres proporciona gran abundancia de información sobre los procesos astronómicos y geológicos que dieron forma al planeta. Todo parece indicar que los cráteres constituyen el vestigio de remotos impactos de enormes meteoritos sobre la superficie de los planetas; tales meteoritos están todavía produciendo cráteres en el actual período geológico. Cual si se tratara de cicatrices producidas por el impacto de cuerpos interplanetarios, los cráteres suministran información sobre la abundancia de tales cuerpos en el espacio interplanetario durante cientos de millones de años. En cuanto rasgos modificados por procesos geológicos, tales como el volcanismo, la actividad tectónica y la erosión, constituyen indicios sobre el medio ambiente que circundaba la superficie de los planetas. Además, si se conoce la velocidad a la cual se originan los cráteres, el simple recuento del número de ellos que se han acumulado sobre una superficie permite averiguar la edad de ésta. Los cráteres constituyen, pues, una clave para reconstruir la historia de cada uno de los planetas interiores.

La fuente histórica constituida por el cráter apenas sí sufre la más leve alteración en la superficie de un cuerpo tal como la Luna, en donde las fuerzas erosivas son del tipo más sencillo. En la superficie de la Luna, sin atmósfera y sin agua, no ha habido sedimentos depositados por el viento en el fondo del cráter, ni olas que hayan limado gradualmente las paredes de éste. Una vez formados los cráteres lunares, es probable que duren por mucho tiempo. Hay, sin embargo, tres maneras principales de borrarlos. En primer lugar, un cráter primitivo puede ser destruido por un segundo impacto o impactos directamente encima de él. En segundo lugar, los cráteres viejos pueden ser recubiertos por materia desprendida en la formación de otros cráteres nuevos. Y, en tercer lugar, los cráteres pueden ser enterrados por ríos de lava.

La manera más directa de obtener información sobre la abundancia de cuerpos interplanetarios que formaron los cráteres es estudiar los cráteres en regiones geológicamente jóvenes, en donde, por estar ampliamente separados, las complejidades de su eliminación resultan despreciables. Al propio tiempo, conviene que los cráteres no se hallen escasamente distribuidos hasta el punto de que se hiciera inseguro el análisis estadístico. En la Luna, las regiones geológicas que más se acercan al ideal para tales estudios son las llanuras de lava que constituyen los "mares". Las muestras de rocas lunares indican que la lava fluyó hace unos 3500 millones de años, barriendo la pizarra de los cráteres preexistentes y creando una superficie limpia para registrar los nuevos. No ha habido actividad interna importante que perturbase los mares desde que se formaron. Durante este tiempo se han acumulado en los mares grandes cráteres, cuyo número suficiente permite constituir una buena muestra que, además, no presenta problemas de superposición u obliteración.

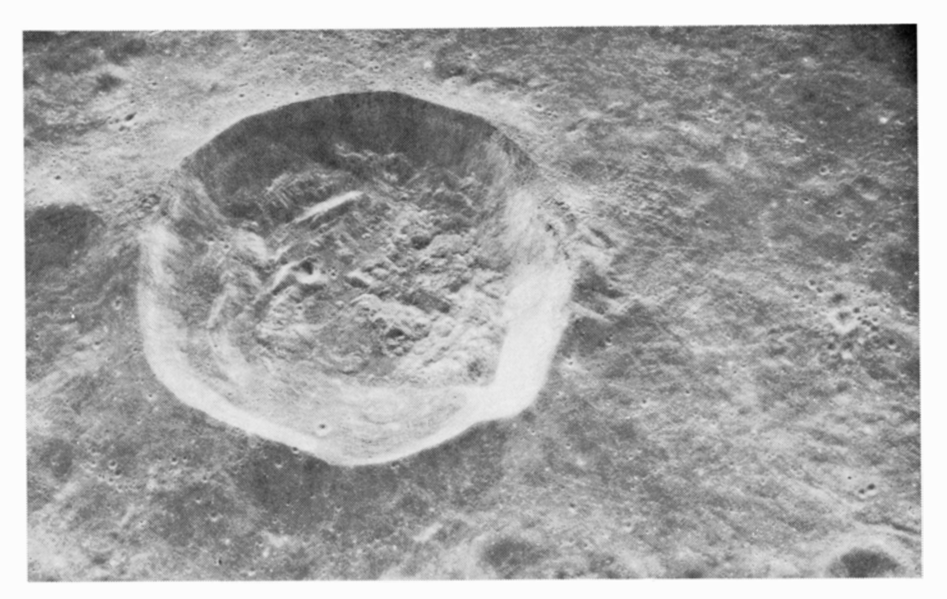
Cuando un meteorito de varios kilómetros de diámetro chocó con la superficie lisa de lava de un mar lunar, los resultados quedaron vívidamente registrados. Una inmensa explosión elevó capas de roca lunar que formaron el borde del cráter. La explosión proyectó hacia fuera chorros de materia a alta velocidad, que cayeron sobre la superficie como largas banderolas de polvo y glóbulos de vidrio: los brillantes "rayos" que se ven mediante telescopios desde la Tierra. La materia expelida a baja velocidad volvió a caer en forma de espesa capa de terreno monticulado que rodea el cráter en una extensión de varias veces su diámetro. La lluvia de materia fue más fuerte en el borde del cráter formado de roca levantada, con lo que aumentó la altura de dicho borde.

Muchos bloques aislados de cascotes volvieron a caer también en el interior, simulando meteoritos y produciendo por sí mismos pequeños cráteres. Hay que distinguir, por tanto, entre la for-

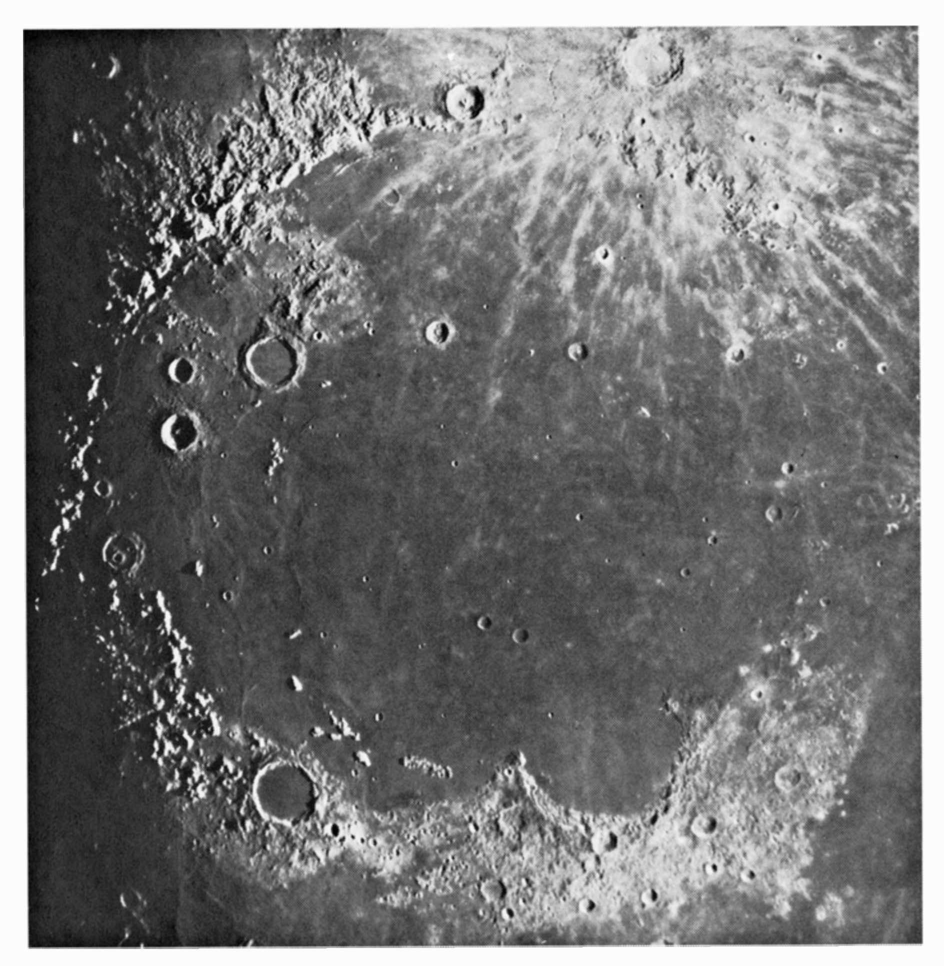


CRATER TERRESTRE de Saskatchewan (Canadá), fotografiado en color convencional por los astronautas del Skylab. Tiene nueve kilómetros de diámetro y cien millones de años de edad. La actividad geológica borró rápidamente la mayoría de los cráteres terrestres, pero muchos cráteres canadienses están en una zona que pudo haberse protegido con los sedimentos eliminados después por los glaciares.

CRATERES MARCIANOS aparecen en este mosaico en color preparado por el Servicio de Inspección Geológica de los Estados Unidos en Flagstaff, Arizona, construido sobre la base de 15 fotografías enviadas por el *Viking I.* Los cráteres, con sus pendientes degradadas por la erosión y la deposición de sedimentos, son los objetos redondeados. Los cañones que penetran en ellos son parte de los Valles Marineris.



CRATERES EN LAS ALTIPLANICIES DE LA LUNA, fotografiados por el *Apolo 10*. El cráter mayor, de unos 40 km de diámetro, es relativamente reciente; su suelo está cubierto de materiales que, al parecer, se han desprendido de las paredes interiores. El terreno circundante está formado por materia expulsada de los cráteres y por cráteres antiguos, parcialmente superpuestos, en diferente estado de degradación. El tamaño de los cráteres pequeños de la fotografía continuará disminuyendo hasta el límite en que ya no sean discernibles.



CRATERES EN LAS TIERRAS BAJAS DE LA LUNA, fotografiados con el telescopio de 100 pulgadas del Monte Wilson. Por estar esta región cerca del Polo Norte de la Luna, la fotografía original mostraba el fondo de la amplia depresión circular que ocupa la mayor parte de la ilustración apreciablemente deformado. Aquí se ve desde la vertical de la depresión porque la imagen de ésta fue proyectada sobre una esfera y luego fotografiada de nuevo desde la vertical del fondo de la depresión. Esta es el Mare Imbrium, la mayor estructura de impacto de la Luna. Se cree formada por un asteroide que tenía entre 100 y 150 km de diámetro. La depresión se fue rellenando luego por ríos de lava, y sobre este nuevo fondo se formaron cráteres más modernos. Arriba, a la derecha, el cráter Copernicus, conspicuo y prominente. Los "rayos" que parten de él es la materia arrojada fuera del cráter por el impacto del cuerpo que lo formó.

mación primaria de cráteres por los meteoritos y la formación secundaria por bloques de materia de la propia superficie. Aunque los cráteres secundarios creados por un impacto son siempre más pequeños que el cráter primario, algunos de ellos llegan a alcanzar unos cuantos kilómetros de diámetro. Por ello, resulta más difícil analizar los cráteres pequeños que los grandes. Además, los cráteres pequeños se pueden confundir con pozos volcánicos enmascarados por desechos de cráteres de impacto próximos. La mayor parte de los cráteres de diámetro superior a cuatro kilómetros están, con toda probabilidad, directamente relacionados con la abundancia de cuerpos meteoríticos en el espacio interplanetario, mientras que la mayor parte de los cráteres menores probablemente no lo están. Consiguientemente, voy a reducir esta discusión a lo que se puede aprender del estudio de los cráteres mayores.

Por ejemplo, de los cráteres mayores se pueden deducir algunas cosas acerca de los propios meteoritos incidentes. Si se representa gráficamente el número de cráteres en un mar lunar en función de su tamaño, se encuentra que el número de cráteres es, aproximadamente, proporcional al cuadrado del inverso de sus diámetros: si existe cierto número de cráteres de diámetro superior a 10 km, una cuarta parte presentará un diámetro mayor que 20 km, una novena lo tendrá mayor que 30 km, y así sucesivamente. De aquí que pueda decirse que los cráteres primarios están distribuidos en un espectro de tamaños que sigue una ley potencial con un exponente igual a –2. Si el gráfico del número de cráteres respecto a su tamaño se construye en papel logarítmico, resulta casi una línea recta con pendiente igual a –2 *(véase la ilustración de la página 74).*

Tal gráfico tiene varias e interesantes consecuencias. En primer lugar, la masa del meteorito requerido para formar un cráter de cada tamaño se puede calcular aproximadamente a partir de impactos de pruebas realizadas en la Tierra. De esos cálculos, el espectro de tamaños de cráteres se puede convertir en un espectro de diámetros de meteoritos. Se encuentra que los diámetros de los meteoritos que produjeron los cráteres están también distribuidos en un espectro que sigue una ley potencial con un exponente aproximadamente igual a –2.

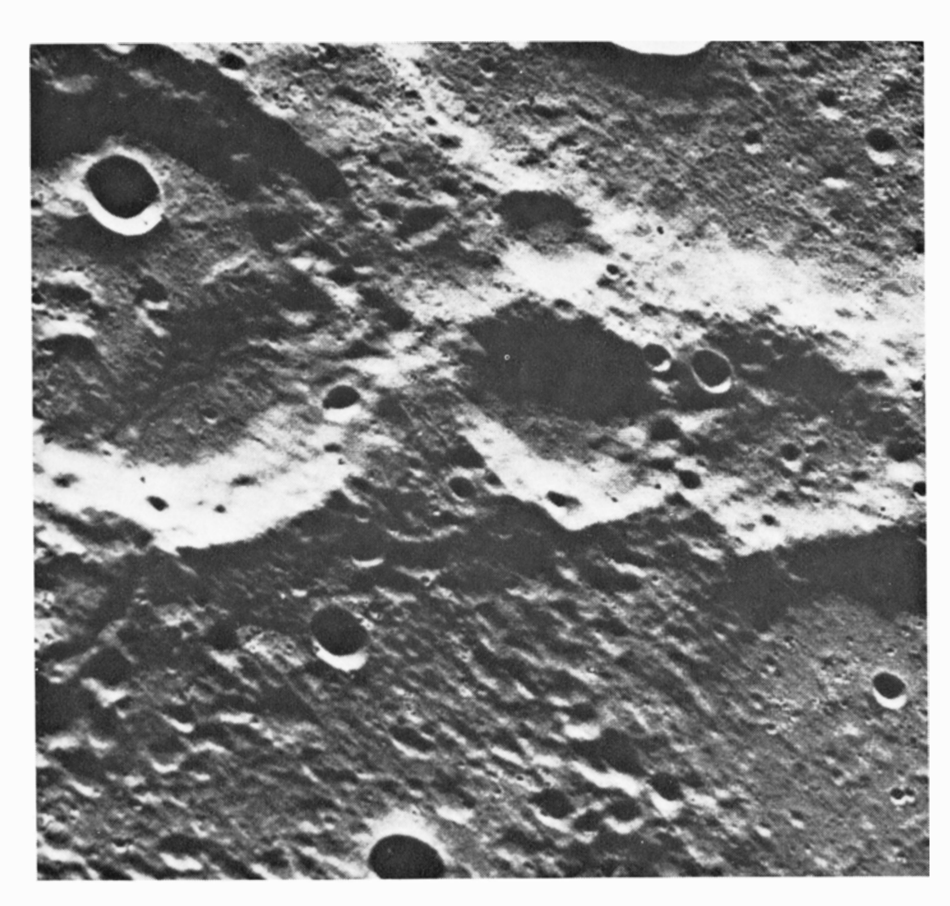
Resulta, pues, que el espectro de tamaños de los meteoritos que formaron los cráteres en la Luna y los planetas es esencialmente igual que el espectro

de tamaños tanto de los asteroides en el cinturón de asteroides como de los meteoritos que caen en la Tierra. Este descubrimiento es una firme confirmación de la hipótesis de que la inmensa mayoría de los cráteres de los planetas y satélites fueron producidos por cuerpos interplanetarios del tipo de los asteroides y los meteoritos. Además, esta clase de espectro de las masas es característica del espectro de las masas de los fragmentos que se producen cuando dos rocas chocan a gran velocidad, es decir, a unos cuantos kilómetros por segundo. Los asteroides chocan a esas velocidades hoy día. Así, uno puede inferir que los cuerpos que produjeron los cráteres eran fragmentos de cuerpos asteroideos que previamente habían chocado y se habían roto en pedazos. Tal deducción concuerda con estudios sobre meteoritos y asteroides, los cuales ponen de relieve la existencia de muchos cuerpos asteroideos en el primitivo sistema solar. De aquí que el estudio de los cráteres de la Luna constituya un eslabón entre estudios astronómicos de cuerpos en el espacio interplanetario y estudios geológicos de la evolución de la corteza de la Luna.

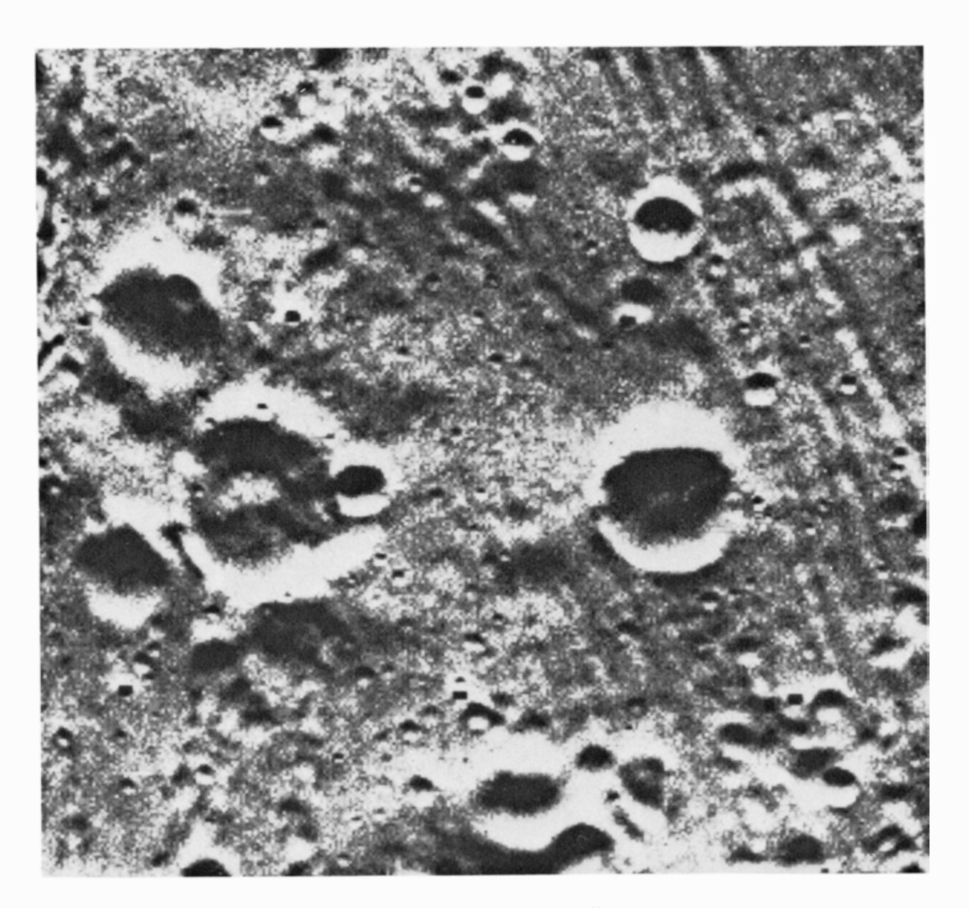
El espectro de tamaños de los cráteres indica además que los meteoritos incidentes deben haber tenido tamaños variables, incluso haber sido cuerpos de más de 100 km de diámetro. Estos cuerpos debieron poseer el aspecto de asteroides. Si se suman todos los cráteres de todos los tamaños que existen en los planetas interiores, se encuentra que el número de meteoritos incidentes hubo de ascender a un porcentaje apreciable del número de cuerpos que hoy viajan en el cinturón de asteroides. Muchos objetos adicionales, sin embargo, pudieron producir cráteres borrados ya por la erosión, ríos de lava, etc. Se estima que para que en todos los planetas interiores se hubieran producido cráteres en el mismo grado que en la Luna, se hubiera necesitado un número aproximadamente igual a la mitad de los cuerpos hoy existentes en el cinturón de asteroides.

Tal número pudo haber representado sólo la última porción de la materia interplanetaria que tomó parte en la formación de los planetas.

Se puede investigar regiones de la superficie de la Luna más viejas, más salpicadas de cráteres y situadas fuera de los mares; la metodología apropiada consiste en combinar estudios de cráteres lunares con análisis de rocas lunares. Los astronautas recogieron muestras de varias provincias geológicas, algunas mucho más antiguas que los mares. Con-



CRATERES DE MERCURIO, fotografiados por el *Mariner 10.* Se asemejan a los que se encuentran en las altiplanicies lunares, hallándose en diferente estado de degradación y superposición.



CRATERES DE PHOBOS la mayor de las dos lunas de Marte, parecidos también a los de las altiplanicies lunares. Esta fotografía fue tomada el día 18 de septiembre a 800 km de distancia por el vehículo orbital *Viking 2.* Los surcos constituyeron un descubrimiento inesperado. Pueden ser fracturas creadas por un impacto en Phobos o, por otro, sobre un hipotético cuerpo padre.

tando el número de cráteres en una provincia cuya antigüedad se conoce a partir de las rocas, cabe determinar la velocidad a la que se produjeron cráteres en la Luna, en diferentes eras, en la historia del sistema solar.

Este enfoque revela que, en época anterior a los últimos 4000 millones de años, la velocidad a la que se producían cráteres en la Luna era de cientos o incluso miles de veces mayor que aquella a la que se producen hoy. Además, la velocidad disminuyó rápidamente desde los tiempos primitivos hasta hace aproximadamente tres mil millones de años, en que alcanzó el ritmo actual. Este se ha mantenido relativamente constante. Se cree que los otros planetas y satélites han sufrido una sucesión análoga.

¿Por qué fueron los planetas tan densamente bombardeados en los primeros tiempos de la historia del sistema solar? Según una hipótesis, la respuesta está directamente relacionada con la manera como los planetas y sus satélites se originaron. Los planetas se formaron cuando la materia en la nube de gases y polvo que rodeaba el primitivo Sol se hubo agrupado gradualmente en grandes cuerpos sólidos. Mientras los planetas y satélites se acumulaban, debieron ser intensamente bombardeados por los restantes cuerpos, grandes y pequeños. Aun cuando el 99 por ciento de la materia primitiva se concentró en los grandes cuerpos, el 1 por ciento restante debió ser conducido, con frecuencia, por el campo gravitatorio de aquéllos y lanzado a órbitas que podrían cortar la órbita de un planeta y chocar con él.

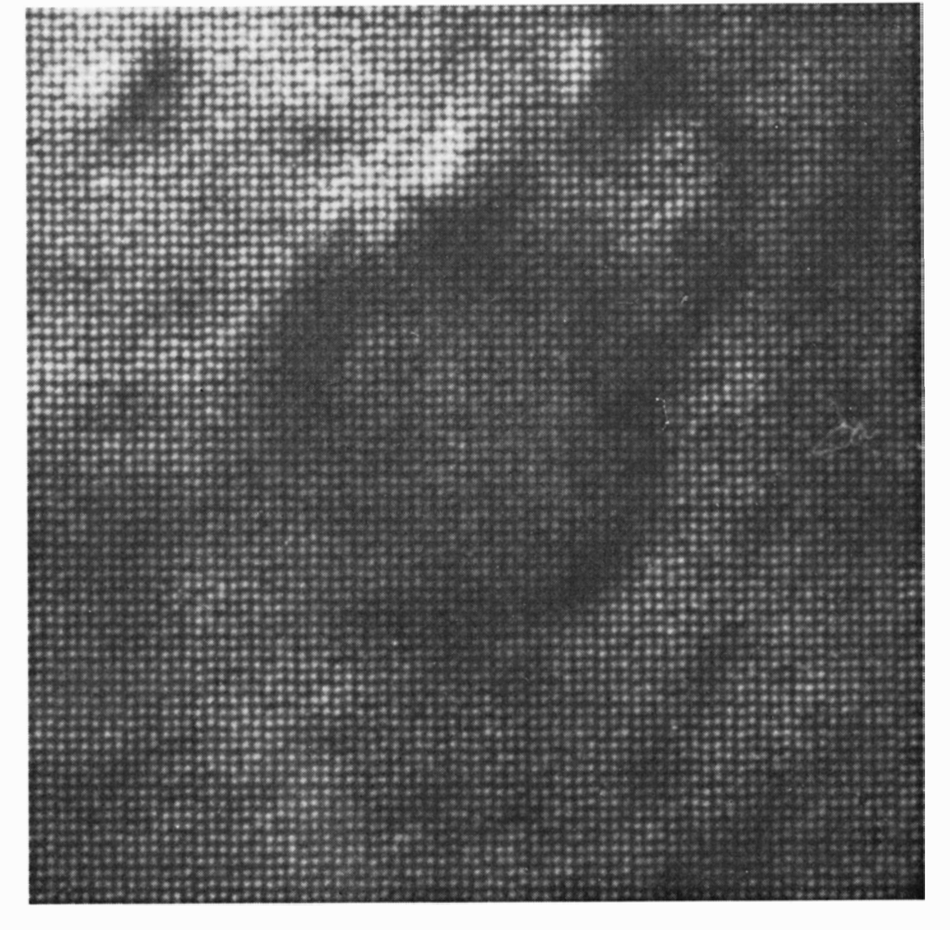
Al parecer, el período durante el cual los planetas fueron intensamente bombardeados fue la última fase del barrido de los restos que quedaron de la primitiva nebulosa. Ciertos cálculos dinámicos muestran que, hace más de cuatro mil millones de años, la vida media de los restos, es decir, el tiempo necesario para que los planetas barriesen la mitad de ellos, pudo haber sido de sólo unos pocos millones de años, como lo indica el rápido decrecimiento de la primitiva producción de cráteres en la Luna. Ese material, por tanto, se habría enrarecido en relativamente poco tiempo. Hoy día, la vida media típica de los cuerpos cuyas trayectorias cortan las órbitas de los planetas está más cerca de 100 millones de años o incluso más. A medida que esos cuerpos son barridos, nuevo material está siendo continuamente suministrado por perturbaciones y colisiones de cuerpos en el "depósito" del sistema solar, a saber, en el cinturón de asteroides.

La determinación de la edad de las rocas lunares ha conducido a una hipótesis distinta sobre la causa que determinó la intensa producción primitiva de cráteres.

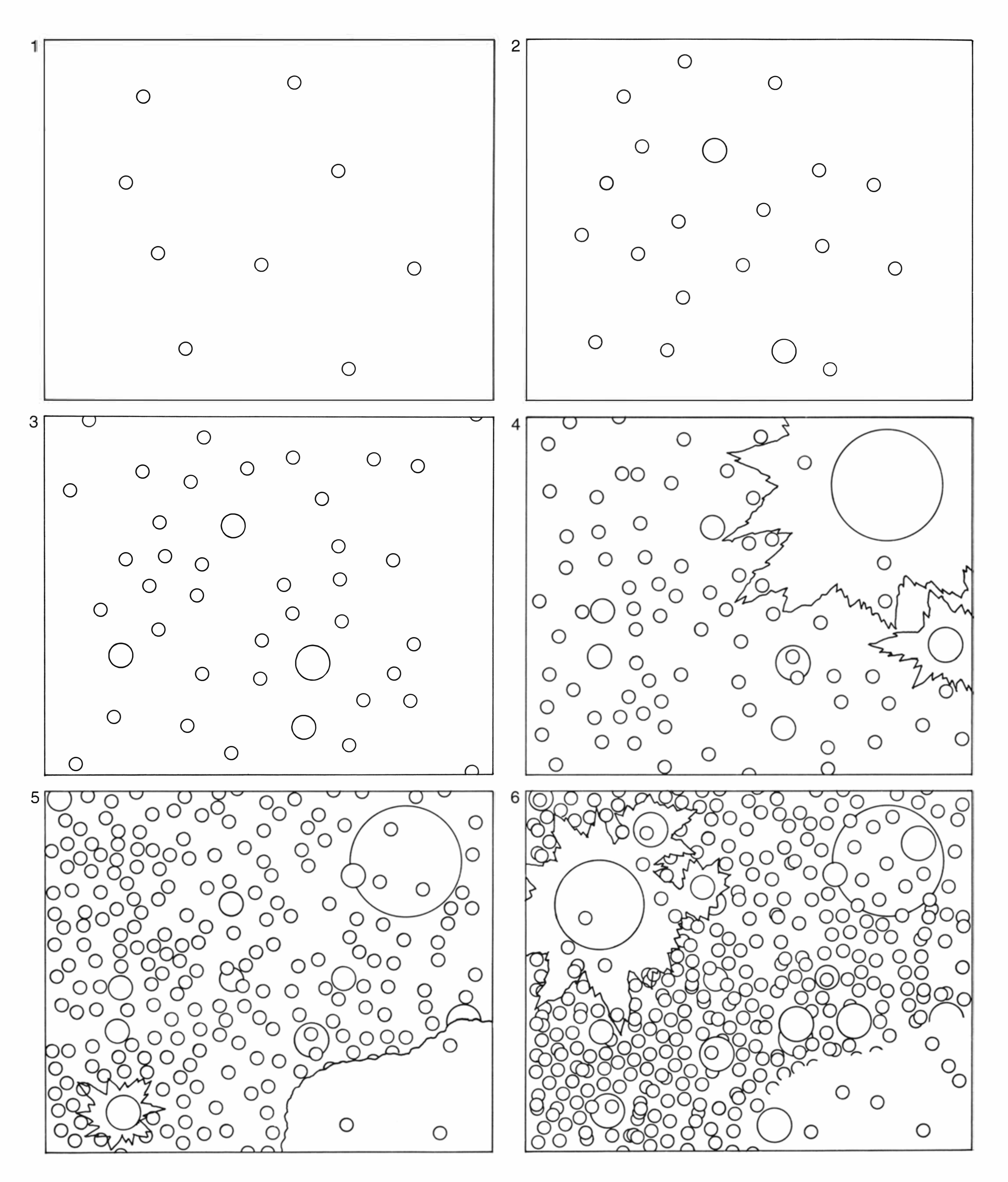
Varios analistas del proyecto Apolo han encontrado que la antigüedad de muchas muestras de rocas lunares se corta bastante bruscamente en los cuatro mil millones de años. Pocas rocas más antiguas han sobrevivido. Gerald F. Wasserburg y sus colaboradores del Instituto de Tecnología de California han sugerido que un breve episodio cataclísmico de intenso bombardeo, que tuvo lugar hace cuatro mil millones de años, destruyó las rocas más antiguas y las estructuras en la superficie del planeta. Tal episodio pudo haber sido el resultado de la rotura en trozos de un gran cuerpo interplanetario que pasó cerca de la Tierra y fue desintegrado por la marea gravitatoria. G. W. Wetherill, del departamento de magnetismo terrestre, de la Carnegie Institution, de Washington, ha calculado que tales cuerpos podrían haber sido desviados desde la región de los planetas gigantes en el sistema solar exterior hacia el sistema solar interno, hace unos cuatro mil millones de años.

La verdad puede encontrarse a mitad del camino entre las dos hipótesis. La curva descendente que describe la velocidad de formación de cráteres casi seguramente muestra "dientes" que representarían explosiones de grupos de fragmentos procedentes de asteroides o cometas desintegrados. Si tales dientes presentaron o no proporciones cataclísmicas, alterando dramáticamente la evolución de la superficie de un planeta, es algo que está por dilucidar.

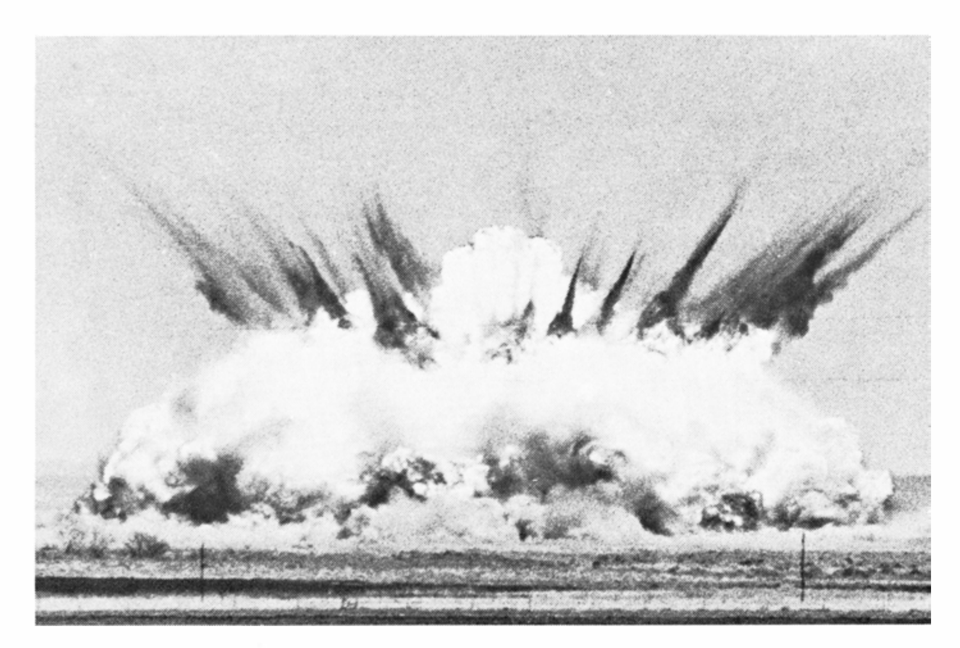
Cualquier modelo de la primitiva estructura geológica de los planetas debe tomar en consideración el capítulo de su historia que habla de su craterización. Por ejemplo, Robert B. Hargraves, de la Universidad de Princeton, y otros han logrado recientes avances en la interpretación de la geología precámbrica de la Tierra al invocar procesos puramente internos para dar cuenta de la rotura de una corteza homogénea hipotética que se hallaría debajo de un mar global. Esta corteza se supone que se formaría entre hace cuatro mil y cuatro



CRATERES DE VENUS, revelados a través de la densa capa nubosa del planeta por una imagen de radar. Aunque este tipo de imagen hace aparecer escabrosos los cráteres, su relieve es muy bajo. El mayor tiene un diámetro de 160 km, pero sus bordes sólo poseen 500 m de altura.



EL NUMERO Y EL TAMAÑO de los cráteres informan sobre los cuerpos que los produjeron. Aquí se muestra la superficie de un planeta hipotético en seis épocas de su historia, a partir de la formación de una superficie reciente *(1)*. Cuando se cuentan y miden los cráteres reales en un área no perturbada de un planeta, el número acumulado de cráteres de cada tamaño concuerda con el número previsto si los cráteres se produjeron por cuerpos incidentes que tuvieran la misma distribución de tamaños que los cuerpos en el cinturón de asteroides. Se puede así inferir que la población de objetos que produjeron los cráteres se parecía mucho a la población de asteroides actual. La distribución de cráteres también suministra información sobre la edad de las diferentes provincias geológicas de un planeta o de un satélite. En provincias jóvenes se encuentran cráteres bien diferenciados de los demás sobre un fondo liso *(primeros tres grabados)*; en provincias viejas, los nuevos cráteres se forman sobre otros más antiguos y tienden a borrarlos *(tres últimos grabados)*. A veces, sin embargo, puede crearse una nueva provincia cuando una capa de materia expulsada de un impacto gigantesco cubre cráteres antiguos *(4)* o una corriente de lava alisa la superficie pizarrosa *(5)*. Pasado un cierto límite, el terreno más antiguo queda casi saturado con cráteres y se acerca al equilibrio. Sin embargo, puede todavía detectarse heterogeneidades en la impresión de los cráteres introducidas por anteriores procesos de obliteración *(6)*.



EXPLOSION DE PRUEBA de 100 toneladas de trinitrotolueno en un desierto de Colorado, que ilustró algunas características de la formación de cráteres. La materia que compone los chorros de gran velocidad en ángulos entre la horizontal y la vertical fue arrojada saltando limpiamente el cráter; ésta es la clase de materia que forma los largos y brillantes rayos que rodean algunos de los modernos grandes cráteres lunares, como el Copérnico. Esta explosión excavó un cráter de unos 40 metros de diámetro y siete de profundidad. Bloques del material expulsado del cráter primario produjeron pequeños cráteres secundarios.

mil quinientos millones de años; su rotura es necesaria para explicar la primitiva evolución continental. La formación de cráteres es un proceso externo con el mismo resultado. En la Luna produjeron depresiones de cientos de kilómetros de diámetro, excavando materiales hasta una profundidad de varios kilómetros al menos y apilándolos en otros lugares, hace más de cuatro mil millones de años. Lo mismo tuvo que ocurrir en la Tierra, dando lugar a una heterogeneidad, a gran escala, en la corteza terrestre.

Muchos, si no todos los planetas, comenzaron su historia con una superficie densamente salpicada de cráteres. Suministran la prueba las antiguas áreas de tierras altas de color claro de la Luna, que forman un laberinto de cráteres superpuestos. Según datos obtenidos de las rocas lunares, la corteza de las tierras altas se constituyó hace entre 4400 y 4500 millones de años, cuando un calentamiento (de origen desconocido) fundió las capas externas de la Luna, creando con el tiempo una superficie de rocas ígneas. Esto ocurrió durante el periodo de intenso bombardeo, de manera que la primitiva corteza quedó rápidamente pulverizada por cráteres que se superponían. Durante los pocos cientos de millones de años inmediatos a la formación de los planetas, se acumularon evidentemente en ellos suficientes cráteres para cubrir no sólo un pequeño porcentaje de su superficie, sino el 100 por ciento de ella. De una tal superficie se dice que está saturada de cráteres. Puesto que en una superficie saturada cada nuevo cráter destruiría, por término medio, el equivalente de otro antiguo del mismo tamaño, la superficie se dice también que está en equilibrio en cuanto a producción de cráteres.

Las consecuencias geológicas de la saturación son importantes. La ley potencial inversa del cuadrado que rige el espectro de distribución de tamaños de cráteres en la Luna tiene la propiedad de que, si hay suficientes cráteres primarios de diámetros entre uno y 10 km para cubrir un planeta entero, también hay suficientes cráteres de diámetros entre 100 y 1000 km para cubrirlo. El impacto de un meteorito suficientemente grande para crear un cráter de 100 km de diámetro excavaría la superficie del planeta hasta una profundidad de varios kilómetros. Así, mientras que un planeta al 10 por ciento de saturación con cráteres de varios kilómetros tendría una superficie consistente en estratos de rocas primarias interrumpidos por cráteres, un planeta totalmente saturado con cráteres tendría una superficie consistente en restos pulverizados de varios kilómetros de espesor. Tal capa superficial se llama megaregolito.

Con toda probabilidad, tal megaregolito se acumuló en la Luna durante sus primeros, pocos, centenares de millones de años. La producción de cráteres fue seguramente suficiente para saturar la superficie, no una, sino muchas veces, ya que presumiblemente la Luna se formó por acumulación de cuerpos que cayeron en ella. Esta conclusión es confirmada por el hecho de que los cráteres lunares se hallan en diversos estados de degradación, que van desde aquellos de aspecto nuevo con bordes agudos hasta los restos erosionados de cráteres viejos, a los que se superponen otros muchos y el material expulsado de ellos. Toda la roca ígnea originaria de la Luna habría sido excavada, pulverizada y redistribuida por la formación de los cráteres. Un corte vertical a través de la corteza mostraría un megaregolito compuesto por capas superpuestas de material expulsado, que se ha hecho compacto, capas que fueron calentadas y parcialmente soldadas entre sí y después fracturadas quizá de nuevo. Tal formación explicaría por qué los astronautas en las tierras altas lunares hallaron no las rocas primarias bien preservadas que algunos geólogos esperaban sino rocas conglomeradas formadas de materiales que habían sido fragmentados y recementados. Podría igualmente explicar algunas de las propiedades sísmicas de la Luna, que descartan la existencia de capas de roca sólida cerca de gran parte de la superficie lunar.

Si no hubiese habido corrientes de lava, actividad tectónica, atmósfera ni océanos en ningún cuerpo del sistema solar interno, la superficie de todos sus planetas tendría el mismo aspecto saturado de cráteres que las altiplanicies lunares. En los planetas más grandes, sin embargo, procesos volcánicos y erosivos, así como la continuada formación de cráteres, han borrado los cráteres antiguos. El número, tamaño y carácter de los cráteres en los diversos planetas interiores es diferente, lo que sugiere diferentes condiciones erosivas, diferentes edades y quizá incluso diferente entorno meteorítico. Cada planeta tiene una formación de cráteres peculiar, que suministra indicaciones sobre su historia.

Las dos lunas de Marte, Phobos y Deimos, son los únicos satélites en el sistema solar interno, aparte de la Luna terrestre. La densidad de cráteres en su superficie es similar a la hallada en las altiplanicies lunares. Se cree que los dos satélites son demasiado pequeños para mostrar alguna actividad interna o superficial que erosionase los cráteres, aunque hay surcos lineales, en Phobos, todavía no explicados. (Podrían ser fracturas originadas por impactos.) Phobos y

Deimos llevan probablemente las cicatrices de miles de millones de años de impactos, quizás en su mayoría debidos a cuerpos errantes del cinturón de asteroides que se encuentra un poco más allá de la órbita de Marte. Ellos mismos pueden ser los fragmentos de un cuerpo mayor que fue partido por un suceso catastrófico de formación de un cráter.

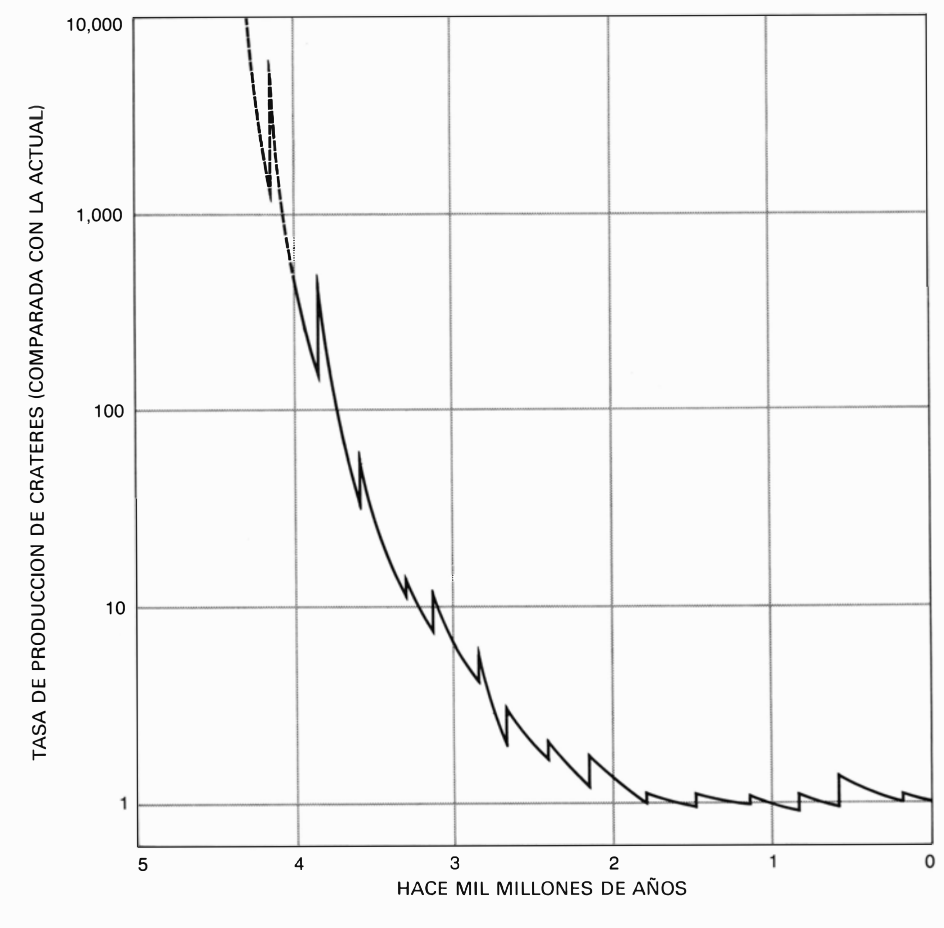
Marte mismo presenta un cuadro muy diferente. El planeta ha sido sometido a intenso estudio desde que el *Mariner 9* suministró mapas fotográficos de la totalidad de su superficie en 1971 y 1972. El descubrimiento más sorprendente es que el planeta está aproximadamente dividido en dos hemisferios, salpicado escasamente uno de cráteres y densamente el otro. En la región escasamente picada hay planicies lisas –al parecer, recientes corrientes de lava– punteadas de cráteres dispersos y de aspecto joven. El análisis del tamaño y el número de los cráteres revela que las condiciones en el planeta Marte son análogas a las halladas en los mares de la Luna. El espectro de tamaños de los cráteres marcianos sigue la conocida distribución potencial con un exponente de aproximadamente –2, que concuerda con la ley hallada para el espectro de masas de los asteroides. Por tanto, los impactos de los asteroides parecen haber sido registrados y preservados en las regiones más jóvenes y menos perturbadas de Marte.

Pero si ensanchamos nuestro objetivo para incluir todo Marte, hallamos fácilmente pruebas de perturbaciones. Las fotografías muestran que existen muchas áreas marcianas, particularmente cerca de los polos, donde se han depositado espesas capas de sedimentos, arrastradas quizá por el viento. En otras zonas, las capas de sedimentos están siendo erosionadas, revelando debajo cráteres antiguos. En el fondo de algunos cráteres se están acumulando dunas de arena. Los canales serpenteantes, o "arroyos", parecen indicar que la superficie del planeta fue en tiempos erosionada por corrientes de agua.

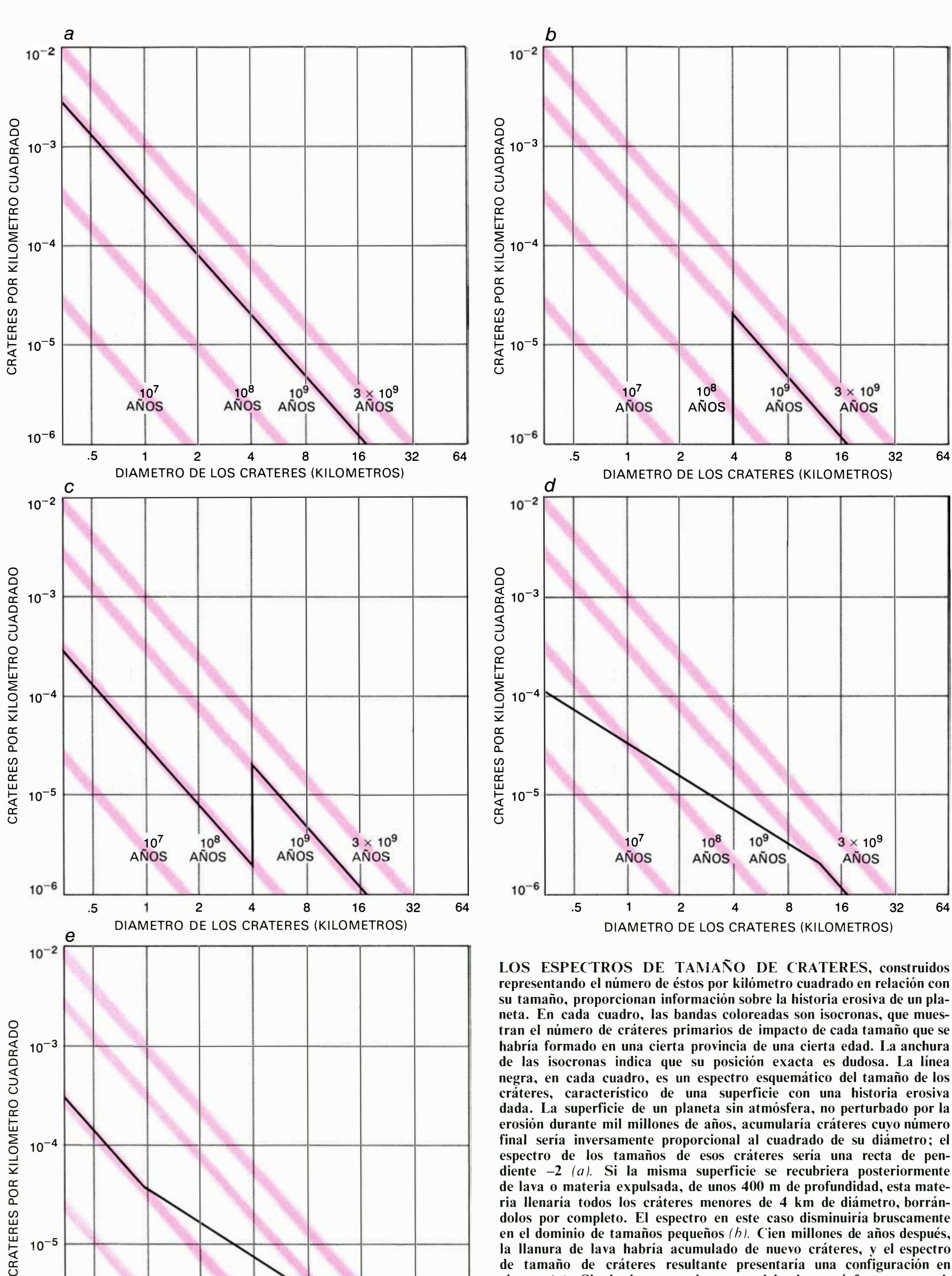
Los estudios de cráteres han constituido la clave para obtener indicaciones sobre los procesos de erosión y deposición en Marte. Si se cuentan los cráteres en las antiguas altiplanicies marcianas, donde muchos cráteres parecen haber sido erosionados o rellenados con sedimentos, se encuentra que su espectro de tamaños difiere marcadamente de la sencilla ley inversa del cuadrado, característica del espectro de tamaños de los cráteres en la mayor parte de las regiones de la Luna o en las llanuras marcianas. Cuando el espectro de tamaños de los cráteres en las altiplanicies marcianas se representa en papel logarítmico, se encuentra un curioso codo, como si la curva se compusiera de dos segmentos que siguieran una ley potencial con un exponente de –2, unidos por un segmento intermedio que siguiera una ley potencial con exponente aproximadamente igual a –1.

Tal codo puede explicarse si Marte hubiera tenido una historia erosiva específica de él. Lo que parece indicar es un largo periodo de erosión y deposición que terminó de manera bastante brusca en época relativamente reciente de la historia marciana. Tal deposición habría llenado los cráteres pequeños y poco profundos, dándoles vidas más cortas que las de los cráteres mayores. Cuanto más grande fuera el cráter, tanto mayor habría sido su probabilidad de supervivencia. El espectro de tamaños de cráteres en régimen estacionario bajo tales condiciones hubiera presentado una pendiente de aproximadamente –1, como lo hace en Marte. Para explicar las observaciones, el dilatado periodo de erosión y deposición debió terminar bruscamente, porque la continuada producción de cráteres ha restablecido la pendiente de –2 para los tamaños pequeños.

En otras palabras, los cráteres marcianos parecen contar una historia de intensa erosión y deposición en alguna época pasada, un periodo que acabó de manera brusca. Numerosos investigadores, incluyendo Raymond E. Arvidson, de la Universidad de Washington, Clark R. Chapman, del Instituto de Ciencia Planetaria, Kenneth J. Jones, de la Universidad Brown, Lawrence A. Soderblom, del Geological Survey (Inspección Geológica) de los Estados Unidos y yo, hemos estudiado cráteres marcianos y confirmado este cuadro general. Los distintos estados de degradación de los cráteres marcianos, que oscilan desde los de aspecto moderno y forma de taza a los llenos de polvo y de fondo plano,



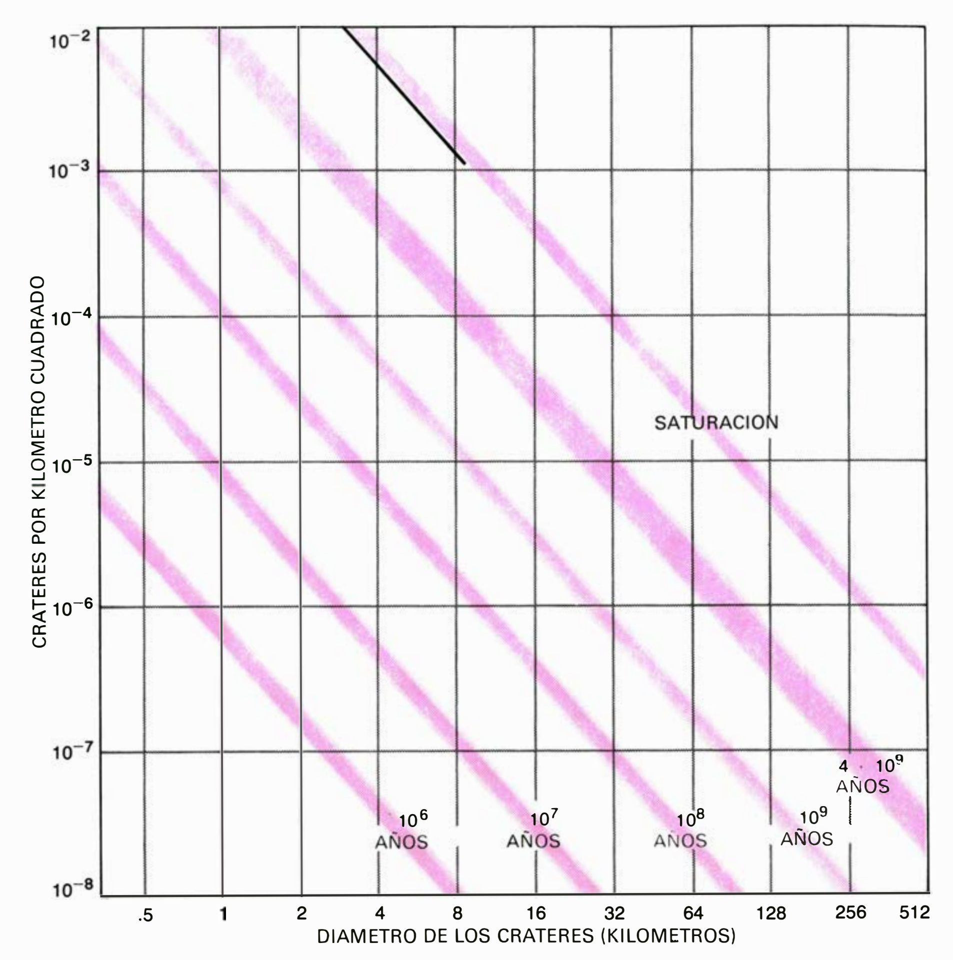
LA VELOCIDAD DE PRODUCCION DE CRATERES ha decrecido durante los últimos 4500 millones de años, de acuerdo con el número de cráteres de la Luna en diferentes provincias, cuyas edades han sido establecidas por análisis de muestras de rocas lunares. La Luna recibió un intenso bombardeo a lo largo de su historia cuando los cuerpos mayores del sistema solar barrían los últimos residuos que quedaron de la formación de dicho sistema. Los máximos agudos de la curva descendente marcan sucesivas épocas (de bombardeo desusadamente intenso) en que el nuevo material estaba siendo inyectado en el sistema solar interno, bien cuando grandes cuerpos chocaron y se rompieron en pedazos o bien cuando el material se desvió del sistema solar externo por el campo gravitatorio de los planetas gigantes.

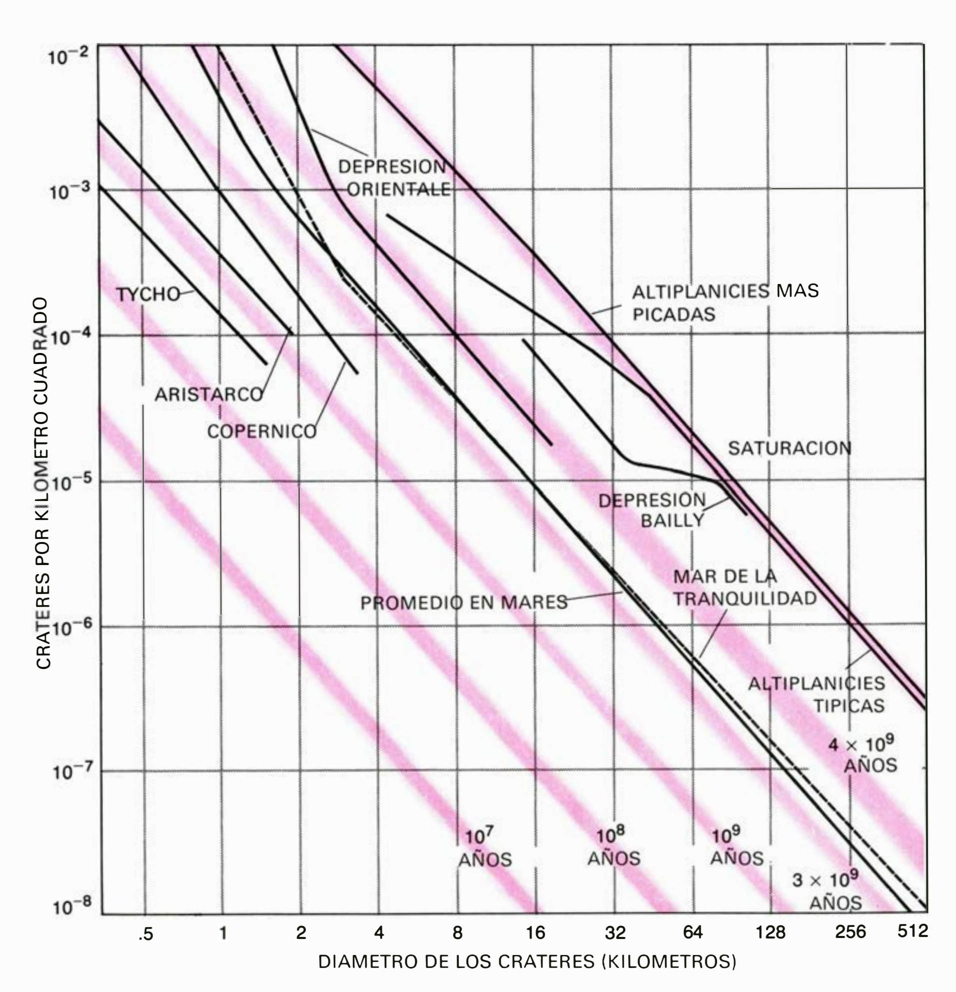


LOS ESPECTROS DE TAMAÑO DE CRATERES, construidos representando el número de éstos por kilómetro cuadrado en relación con su tamaño, proporcionan información sobre la historia erosiva de un planeta. En cada cuadro, las bandas coloreadas son isocronas, que muestran el número de cráteres primarios de impacto de cada tamaño que se habría formado en una cierta provincia de una cierta edad. La anchura de las isocronas indica que su posición exacta es dudosa. La línea negra, en cada cuadro, es un espectro esquemático del tamaño de los cráteres, característico de una superficie con una historia erosiva dada. La superficie de un planeta sin atmósfera, no perturbado por la erosión durante mil millones de años, acumularía cráteres cuyo número final sería inversamente proporcional al cuadrado de su diámetro; el espectro de los tamaños de esos cráteres sería una recta de pendiente −2 *(a)*. Si la misma superficie se recubriera posteriormente de lava o materia expulsada, de unos 400 m de profundidad, esta materia llenaría todos los cráteres menores de 4 km de diámetro, borrándolos por completo. El espectro en este caso disminuiría bruscamente en el dominio de tamaños pequeños *(b)*. Cien millones de años después, la llanura de lava habría acumulado de nuevo cráteres, y el espectro de tamaño de cráteres resultante presentaría una configuración en zigzag *(c)*. Si el planeta tuviera una delgada atmósfera que continuamente transportara sedimentos dentro de los cráteres, los pequeños se borrarían antes que los mayores. Los cráteres pequeños tendrían una pendiente más suave, de aproximadamente −1 *(d)*. Si hace 100 millones de años la deposición de sedimentos hubiera cesado bruscamente, los cráteres pequeños acumulados con posterioridad no se habrían erosionado y su distribución recuperaría la pendiente de −2 *(e)*.

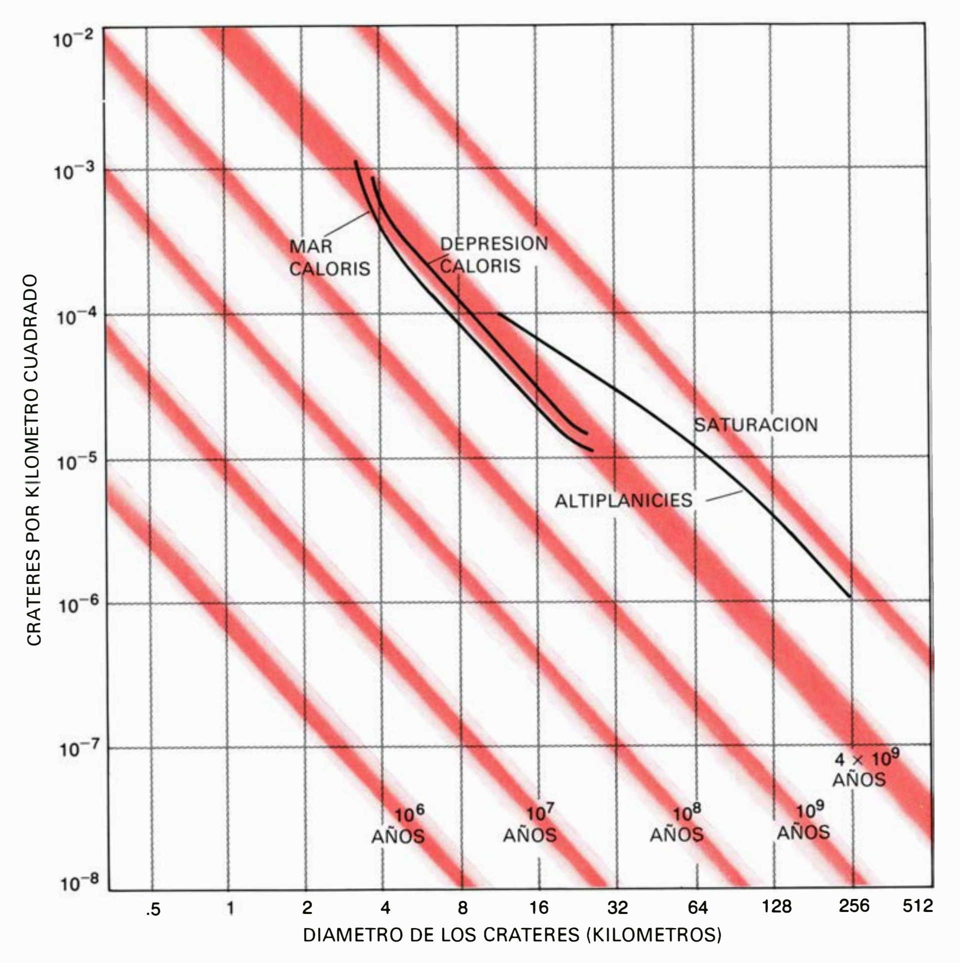
son compatibles con un periodo de intensa erosión. Hay algún desacuerdo, sin embargo, sobre si hubo un solo periodo de erosión o varios y sobre si la mayor parte de la erosión es antigua o relativamente reciente. Una erosión muy antigua concordaría con las indicaciones, recientemente puestas de manifiesto por la nave espacial *Viking*, de que en los primeros tiempos la atmósfera de Marte tenía mayor masa y era por tanto mas densa de lo que es ahora. Los episodios recientes de erosión, si es que han tenido lugar, podrían significar que el clima marciano presenta fluctuaciones como el clima de la Tierra, con sus periodos glaciares y las épocas cálidas interpuestas. Michael C. Malin, del Instituto de Tecnología de California, ha hecho uso de los recuentos de cráteres para estimar que los canales marcianos, que pueden ser reliquias de los periodos erosivos, tienen una antigüedad de miles de millones de años. Si es así, ello apoyaría la hipótesis de que la erosión estuvo asociada con una atmósfera primitiva densa.

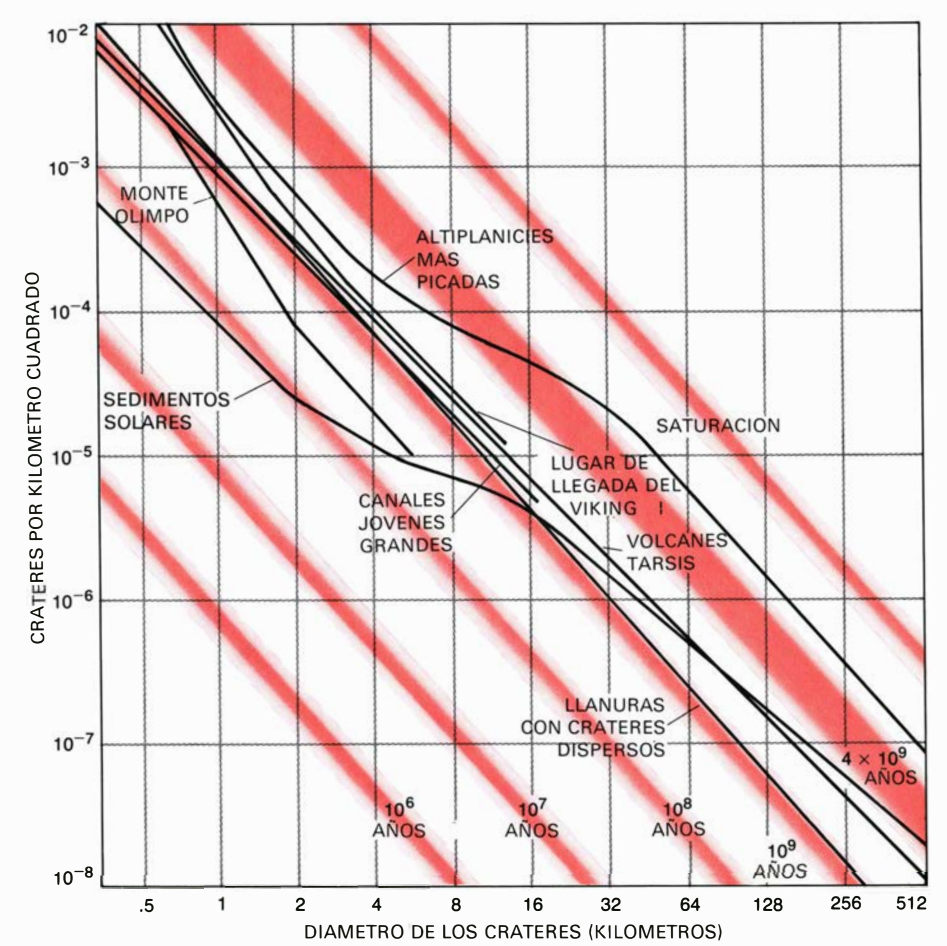
En la Tierra, por supuesto, la erosión es mucho más intensa que en ninguno de los otros planetas interiores. La Tierra es tan activa que, en términos geológicos, los cráteres duran sólo muy poco tiempo. La mayor parte de los cráteres terrestres que sobreviven han sufrido una intensa erosión, y muchos de ellos han sido reconocidos por los geólogos sólo hace pocos años. Una provincia geológica donde la búsqueda de cráteres ha sido particularmente fructífera es el escudo canadiense, la región en torno a la bahía





LOS ESPECTROS DE TAMAÑOS DE CRATERES en los satélites Phobos y Deimos *(arriba)* **y en la Luna** *(abajo)* **muestran que los cuerpos menores del sistema solar interno han sido poco modificados por la erosión. Las isocronas** *(en color)* **representan el número de cráteres primarios que, en teoría, se habrían formado en provincias de diferentes edades. La banda rotulada "Saturación" indica el espectro de tamaños de cráteres que podría esperarse en un cuerpo cuya superficie está saturada de cráteres. Las líneas negras continuas representan el número de cráteres observado en diferentes provincias de los satélites marcianos y la Luna. Phobos y Deimos aparecen casi saturados, luego su superficie es muy antigua. La Luna, sin embargo, presenta claramente provincias de distintas edades. La curva rotulada "Altiplanicies típicas" indica que algunos cráteres pequeños en algunas de las altiplanicies han sido probablemente borrados por lava. Las curvas desusadamente escarpadas en la parte alta del espectro indican que, en ciertas áreas, se han incluido probablemente cráteres secundarios de impacto y cráteres volcánicos en el recuento de cráteres primarios. La curva en zigzag para la Depresión Bailly puede indicar que algunos cráteres menores en la depresión se borraron en alguna época por productos expulsados en choques.**





de Hudson, con formaciones estables de alrededor de mil millones de años. En esos mil millones de años, que constituyen casi un tercio de la vida de los mares lunares, el impacto de los meteoritos ha excavado muchos cráteres. Los más pequeños fueron rápidamente borrados. Los mayores, de varias decenas de kilómetros de diámetro, fueron aplanados por glaciares y parcialmente rellenados por sedimentos. Algunos están siendo exhumados por la erosión de los antiguos sedimentos, y sus restos han sido identificados en número creciente por geólogos canadienses; entre ellos C. Beals y Michael R. Dence y sus colaboradores. Se han encontrado características semejantes en provincias geológicas más jóvenes, incluyendo zonas no montañosas de los Estados Unidos.

Si el número de cráteres por kilómetro cuadrado de las provincias terrestres con tales formaciones se divide por la antigüedad media de las provincias, se puede hallar la velocidad a la cual se han ido produciendo los cráteres durante los últimos mil millones de años. La velocidad de producción de cráteres en el escudo canadiense concuerda aproximadamente con la velocidad estimada para los últimos tres mil millones de años en las llanuras lunares. Algunos investigadores, al comparar las dos velocidades, han sugerido que la velocidad media en los últimos mil millones de años ha sido algo más alta que el promedio para el total de tres mil millones. Si ello es así, significaría que las velocidades varían con el tiempo a medida que los asteroides o cometas aportan nuevos fragmentos de meteoritos al sistema solar interno.

El propio escudo canadiense, relativamente estable, revela con claridad los efectos de la erosión sobre la población de cráteres, lo que permite confirmar la validez de la técnica de deducir procesos erosivos partiendo de las peculiari-

LOS ESPECTROS DE TAMAÑOS de cráteres, de Mercurio y Marte, indican que las superficies de esos cuerpos mayores han sido más profundamente alteradas con el tiempo. El recuento de cráteres en ciertas áreas de muestra en Mercurio *(arriba)* indica que la mayor parte de la superficie de Mercurio es antigua. El déficit de pequeños cráteres en las altiplanicies de Mercurio puede indicar que la superficie del planeta se hallaba en un estado plástico cuando fue intensamente bombardeada, en cuyo caso los cráteres menores habrían desaparecido. Las curvas del espectro para Marte *(abajo)* muestran que las más antiguas de las provincias con cráteres experimentaron un intenso periodo de erosión y deposición que acabó hace algunos miles de millones de años. Los volcanes y canales marcianos son más modernos. Las regiones polares parecen haber sufrido un periodo posterior de erosión y deposición que pudo haber terminado hace unos 100 millones de años.

dades del espectro de tamaños de los cráteres. Los cráteres pequeños del escudo canadiense han sido borrados tan rápidamente que el espectro de tamaños presenta un fuerte déficit de cráteres pequeños. Así, un observador de otro planeta que sólo dispusiera de las fotografías tomadas por un vehículo espacial, llegaría a la conclusión de que la mayor parte de la superficie de la Tierra se formó en épocas geológicas recientes (basándose en la escasez general de cráteres de impacto) y que los procesos de erosión y deposición están activamente destruyendo, en nuestros días, las estructuras geológicas pequeñas (basándose en el déficit de cráteres pequeños en particular). Por ejemplo, las estadísticas de cráteres deducidas de diversas regiones en la Tierra indican que un cráter terrestre de 100 metros de diámetro tiene una vida típica de unos pocos miles de años, otro cráter de un kilómetro de diámetro tiene una vida de un millón de años, otro de 10 km de diámetro una vida de unos cuantos cientos de millones de años y otro de 100 km de diámetro tiene una vida de algunos miles de millones de años.

Venus parece presentar un cuadro intermedio entre el de Marte y el de la Tierra. Richard M. Goldstein y sus colaboradores del Laboratorio de propulsión a chorro del Instituto de Tecnología de California obtuvieron, mediante radar, imágenes de la superficie de Venus, oculta por densas capas nubosas; y las imágenes muestran cráteres. Sólo una pequeña fracción del planeta ha sido cartografiada, pero el recuerdo preliminar de cráteres da dos impresiones. En primer lugar, los cráteres de diámetros mayores que 100 o 200 km son casi tan abundantes en Venus como lo son en las regiones más antiguas de Marte, aunque son menos abundantes en Venus que en las altiplanicies lunares. Resultan unas diez veces más abundantes en Venus que en el escudo canadiense, de la Tierra. En segundo lugar, la población de cráteres de menos de 100 km de diámetro ha sido fuertemente diezmada, probablemente por la erosión. La hipótesis de la erosión viene apoyada por el descubrimiento, hecho mediante el radar, de que los cráteres son muy poco profundos y pueden estar parcialmente rellenos de polvo.

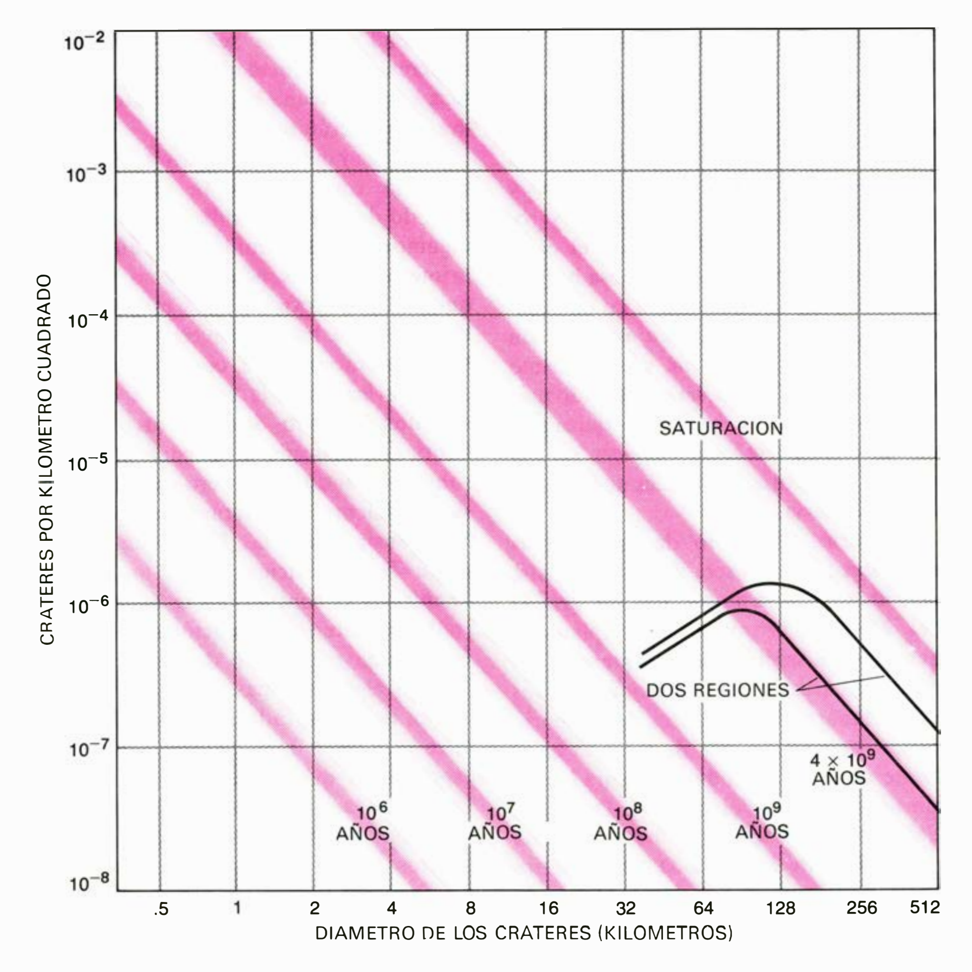
Se puede inferir que la superficie de Venus no es tan vieja como la superficie de la Luna ni tan activa geológicamente como la de la Tierra. La superficie actual de Venus debe haberse formado en una época intermedia en el pasado, y la erosión puede estar deshaciendo pequeños cráteres hoy. Esta deducción concuerda con las fotografías enviadas por los vehículos rusos que se posaron en la superficie de Venus en 1975. Las fotografías muestran rocas agudas y angulosas en uno de los lugares y rocas alisadas y redondeadas en el otro, implicando que la materia rocosa ha sido a la vez creada y erosionada en un pasado geológico reciente. La implicación de que Venus, que tiene casi el mismo tamaño que la Tierra, ha preservado más cráteres antiguos y ha tenido quizá menos actividad tectónica requiere ser explicada por teorías generales de la evolución planetaria.

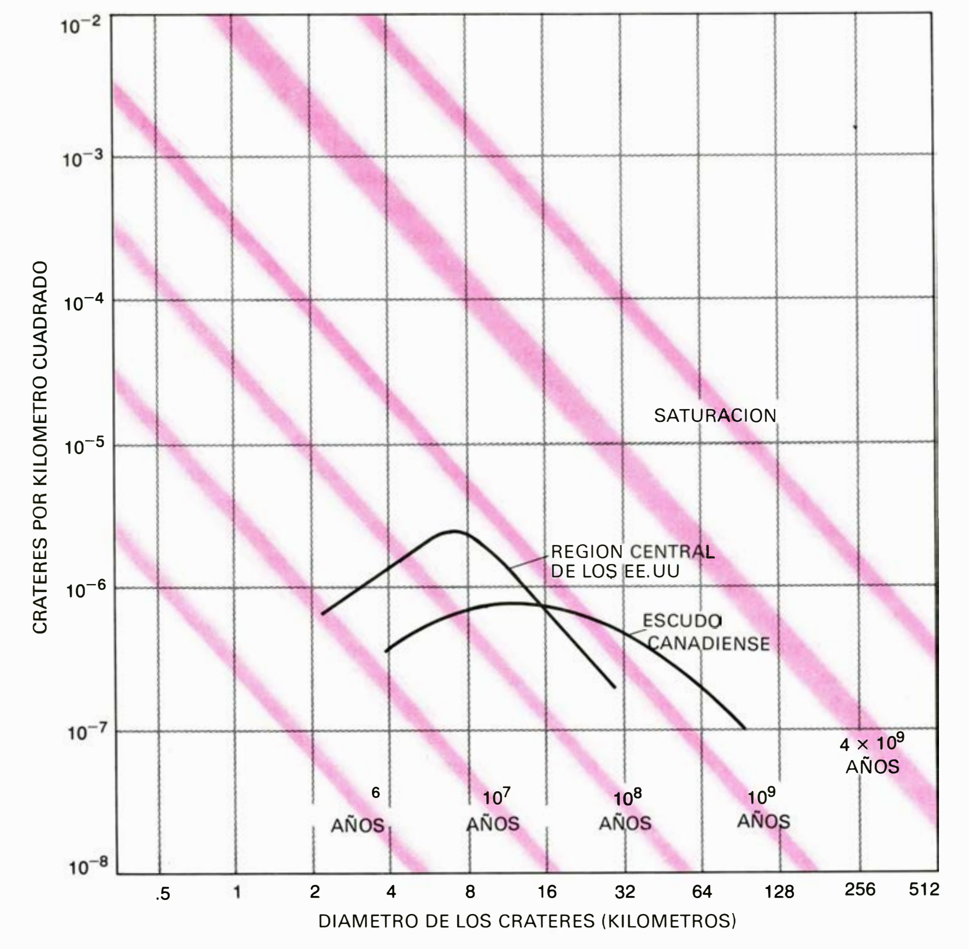
Mercurio, el último de los planetas interiores, parece ser, a primera vista, casi idéntico a la Luna. Las fotografías del *Mariner 10* muestran altiplanicies densamente salpicadas de cráteres y llanuras oscuras y lisas semejantes a los mares lunares. Los cráteres de Mercurio varían desde embudos de tamaño en el límite de resolución de la cámara (aproximadamente un kilómetro) a la Depresión Caloris, de 1300 km de diámetro rodeada por anillos de fractura de una anchura que ronda los 2000 km. Un examen más detenido revela, sin embargo, que el planeta difiere de la Luna en tres importantes aspectos. Primero, la superficie de Mercurio presenta fallas de compresión de cientos de kilómetros de longitud, lo que parece implicar que la totalidad de la corteza del planeta ha contraído su perímetro en una cuantía de unos dos kilómetros. Segundo, incluso en algunas de las regiones de gran densidad de cráteres, la superficie no está saturada de ellos; por el contrario, los cráteres parecen haber perforado una llanura de fondo relativamente lisa. Tercero, Mercurio muestra un leve déficit de cráteres pequeños con respecto al número que se podría esperar si los cráteres se hubieran acumulado al mismo ritmo con que lo han hecho en la Luna.

Entre las interpretaciones que se dan a esas observaciones, hay una que guarda relación con el hecho de que la desintegración radiactiva en el interior de un planeta cause que éste se caliente a medida que transcurre el tiempo. Puede ocurrir que, al ser Mercurio mayor que la Luna, se haya calentado más lentamente y que su corteza se haya enfriado más tarde, en el periodo de intenso bombardeo. Por tanto, la configuración final de la superficie de Mercurio, es decir, si iba a quedar dominada por los cráteres o por una superficie fundida lisa, habría dependido de la lid entre la velocidad a la que se iban formando cráteres y la velocidad a la que la corteza se solidificaba hasta el punto de poder retener cráteres.

Supongamos que Mercurio tenía todavía una corteza fundida al final del periodo de intenso bombardeo primario, digamos hace 4100 millones de años. Los primeros cráteres pequeños no habrían quedado preservados. Los mayores habrían cedido hasta convertirse en sombras, isostáticamente compensadas, de su ser original. Si la corteza se hubiera solidificado poco después, los cráteres habrían empezado a quedar impresos sobre el fondo liso, pero no hubieran sido suficientes para saturar la superficie. Por consiguiente, las fallas que fracturaron la superficie cuando la corteza se enfrió y se contrajo no se habrían borrado por ulterior formación de cráteres. Una fusión residual, o delgados ríos de lava en algunas zonas, podrían, sin embargo, haber borrado algunos de los cráteres primitivos más pequeños. Los tres aspectos en que Mercurio difiere de la Luna quedarían así explicados. Robert G. Strom, de la Universidad de Arizona, y otros investigadores han captado indicios de esos procesos en la Luna, donde no quedan tan de manifiesto porque la corteza lunar se solidificó antes y, por tanto, quedó más afectada por los impactos y la formación de cráteres.

La adición de Mercurio a la serie de superficies planetarias estudiadas ha obligado a los geólogos planetarios a discusiones para interpretar la primitiva historia de la formación de cráteres en los planetas, y tales discusiones están sobrepasando los límites de los hechos conocidos. Hay muchas explicaciones diferentes de la que he ofrecido aquí. Por ejemplo, algunos analistas de los datos del *Mariner 10* arguyen que el mismo episodio de formación cataclísmica de cráteres supuesto para la Luna afectó a Mercurio, llegando después que la superficie lisa de Mercurio se había ya formado y dando nacimiento a la mayor parte de los cráteres visibles en un espacio de tiempo muy corto. Otros investigadores, comparando Mercurio y otros planetas, han propuesto que el déficit en el número de pequeños cráteres proviene no de que los pequeños cráteres fueran borrados, sino de que nunca se llegaron a formar. Razonan que la falta de pequeños cráteres refleja diferencias reales en la distribución de tamaños de los meteoritos incidentes en distintas eras y lugares del sistema solar. Incluso las tradicionales interpretaciones de equilibrio en la producción de cráteres y la saturación de una superficie



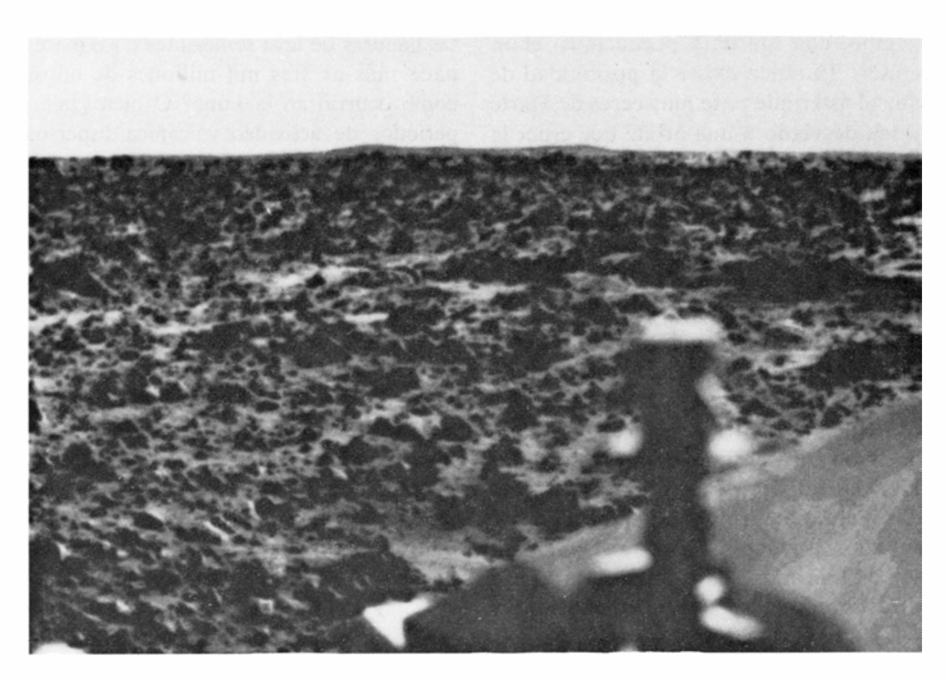


planetaria han sido puestas en duda. Esas dificultades quedan por resolver.

La formación de cráteres en los planetas interiores y sus satélites y lo que los cráteres nos dicen sobre las condiciones geológicas en la superficie de esos cuerpos se puede resumir en el orden de la masa de los cuerpos. Phobos y Deimos son mundos menudos y están completamente cubiertos de cráteres. La Luna está densamente salpicada de ellos, con sólo un 15 por ciento de su más antigua superficie borrada por recientes ríos de lava que representan los mares. Mercurio es muy parecido a la Luna, pero su actividad térmica interna duró más y dejó más huellas. La mitad de la antigua superficie de Marte ha sido modificada por el volcanismo, y hendiduras semejantes a cañones y fracturas hacen pensar que el planeta ha experimentado alguna actividad tectónica rudimentaria. Venus tiene grandes cráteres dispersos, pero el testimonio del radar en cuanto a cañones y rellenado de cráteres sugiere que está en la actualidad siendo modificado por la actividad erosiva y tectónica. La Tierra es tan activa que ya no tiene cráteres de 4000 millones de años de antigüedad y sólo unos pocos, fuertemente erosionados, de antigüedad del orden de 1000 millones de años. Parece existir una tendencia en el sentido de que cuanto mayor es un planeta, mayor es la energía interna que genera y mayor el tiempo durante el cual mantiene procesos tales como el volcanismo y la actividad tectónica. Podría ser que solamente la Tierra tenga tectónica de placas plenamente desarrollada, con corrientes térmicas en el manto empujando las placas una contra otra y arrugando la corteza en cordilleras diferentes de las halladas en cualquier otro planeta. Aunque los cráteres de la Tierra parecen ser sólo cicatrices superficiales, confirman una secuencia fundamental de la evolución planetaria que podría ser menos evidente si hubieran estado ausentes.

LOS ESPECTROS DE TAMAÑOS de cráteres, de Venus y la Tierra, muestran que los dos mayores planetas interiores son geofísicamente los más activos. El espectro para Venus *(arriba)* indica que los cráteres de 32 km de diámetro pueden haber sobrevivido solamente los últimos mil millones de años, mientras que los cráteres de diámetro superior a 128 km podrían datar de hace 4000 millones de años. Los cráteres menores de 32 km no aparecen porque las imágenes de radar de Venus no tienen resolución suficiente para poder detectarlos. En el espectro de tamaños de cráteres de la Tierra *(abajo)* se ve que el número de cráteres que han sobrevivido es función de su diámetro. Ni siquiera los muy grandes lograron sobrevivir desde los primeros días de la historia terrestre.

Hay un efecto adicional relacionado con los impactos que podría ser de importancia fundamental en la determinación del carácter de un planeta. Este efecto está ligado al número, estadísticamente pequeño, de cuerpos muy grandes que incidieron sobre cada planeta en los tiempos primitivos. Cualquier población de meteoritos que origine cráteres en un planeta constará de muchos cuerpos pequeños, unos de tamaño mediano y otro único y muy grande. Esto es así por la naturaleza de los procesos de fragmentación que rigen la distribución de tamaño de asteroides y meteoritos. Las mayores depresiones de impacto en la Luna, Mercurio y Marte tienen unos 1000 km de diámetro, y el de los cuerpos que lo produjeron debe haber sido de unos 100 km. Sin embargo, como hemos visto, los cráteres observados en los planetas pueden ser sólo los últimos de una serie que se extiende en el tiempo hasta los comienzos del sistema solar. Los primeros cráteres pueden no haber sobrevivido en absoluto.



FOTOGRAFIA DEL SUELO DE MARTE, hecha por el *Viking I*, que se posó allí. En el horizonte, una cresta parecida al perfil del Cráter Meteoro. Puede tratarse de un cráter de impacto.

Hay buenas razones para creer, basándose en extrapolaciones de la densidad de cráteres en la Luna junto con distribuciones, estimadas teóricamente, de los tamaños de los cuerpos incidentes y sus órbitas, que los mayores cuerpos que chocaron con cada planeta tenían diámetros de 1000 kilómetros o más. Ahora bien, un asteroide de 1000 km de diámetro que chocase con un planeta primitivo podría haber dado lugar a una asimetría fundamental en el planeta, quizá desprendiendo la corteza de un lado. Aun cuando las huellas de este suceso cataclísmico podrían haberse borrado por formación de cráteres o por procesos tectónicos, la asimetría subyacente podría haberse mantenido en la geología del planeta. Tal colisión podría explicar la asimetría hemisférica de la Luna, cuya corteza en la cara frontal a la Tierra es más delgada que en la cara opuesta. La misma clase de colisión puede ser la causa de la asimetría de Marte, donde un hemisferio tiene muchos cráteres antiguos y el otro ha sido casi enteramente modificado por el volcanismo. Podría, finalmente, haber originado la primitiva rotura de la corteza terrestre en trozos delgados y trozos gruesos.

El impacto de grandes asteroides puede explicar bien el hecho de que el carácter y evolución de los planetas son menos uniformes de lo que podría esperarse si los planetas se hubieran formado mediante una evolución pura, por soldadura de miriadas de partículas pequeñas. Los cuerpos interplanetarios mayores llevaban probablemente tanta energía cinética y cantidad de movimiento que, dependiendo de la dirección desde la cual incidieran sobre un planeta, podrían haberlo inclinado, aumentado o disminuido su velocidad de rotación, destruido un satélite o quizá incluso haber dejado anillos de material en su entorno, tras desintegrarse bajo la acción de las fuerzas gravitatorias.

Hasta ahora he descrito como la formación de cráteres puede suministrar indicaciones tanto sobre el espacio circundante en el que se formaron los planetas como sobre los procesos superficiales e internos que los han modificado. El estudio de los cráteres puede contribuir también a determinar la edad de la superficie de un planeta. En un planeta dado, las antigüedades relativas aparecen claramente, ya que las áreas densamente pobladas de cráteres son más antiguas que aquellas otras que lo están raramente. Es posible estimar también la edad absoluta. Si se conoce la velocidad de producción de cráteres durante cada mil millones de años de la historia de un planeta, y si se cuenta el número de cráteres existentes ahora por kilómetro cuadrado en la superficie de diferentes provincias con diferentes historias erosivas, se puede calcular directamente la edad absoluta de esas provincias en años: basta dividir el número de cráteres por la velocidad de producción de ellos. Esta edad de retención de cráteres difiere en carácter de la edad determinada midiendo la desintegración de isótopos radiactivos en rocas superficiales. Esta última se refiere a sucesos específicos en la historia de ciertas rocas, tales como origen por solidificación o metamorfismo. La edad de retención de cráteres se refiere al tiempo durante el cual una superficie es capaz de retener una estructura topográfica de tamaño dado. La edad de retención de cráteres depende del diámetro del cráter, ya que uno pequeño es fácil que sea erosionado o borrado más de prisa que otro mayor. En resumen, las edades de retención de cráteres miden los niveles de actividad geológica.

Si se conociera la velocidad exacta de producción de cráteres para un planeta, se podría calcular la edad de retención de cráteres para cada provincia del planeta, ya que las provincias densamente salpicadas de cráteres son más antiguas que las que lo están ligeramente. El problema es que la velocidad de producción de cráteres pudo haber sido diferente en distintas partes del sistema solar. El simple hecho de que tengamos medidas absolutas de la velocidad de producción de cráteres en la Tierra y en la Luna no da automáticamente la velocidad de producción de cráteres para otros planetas. Por tanto, carecemos de cronologías absolutas para Mercurio, Venus, Marte, Phobos y Deimos.

La mejor manera de abordar el problema de la datación absoluta es calcular, para los asteroides y cometas conocidos, el ritmo al cual chocan ahora con otros planetas. Este ritmo resulta ser altamente dependiente de la órbita de esos cuerpos. Por ejemplo, un asteroide cuya órbita se extienda desde el cinturón de asteroides hasta un punto inmediatamente interior de la órbita de Marte, puede tener una elevada probabilidad de chocar, con el

tiempo, con Marte, y producir en él un cráter. También existe la posibilidad de que el asteroide pase muy cerca de Marte y sea desviado a una órbita que cruce la de la Tierra, de modo que llegue a producir un cráter en la Tierra o en la Luna. En otras palabras, la órbita inicial del cuerpo tiene una acusada influencia sobre la probabilidad de que llegue a chocar con otro planeta. Wetherill y otros han usado este enfoque probabilístico y han hallado que la velocidad de producción de cráteres en cualquiera de los planetas o satélites interiores no difiere de la de otro en más de un factor diez. Las velocidades pueden llegar a concordar incluso dentro de un factor tres. Resultados análogos han sido recientemente comunicados como fruto de un censo de asteroides por Eugene M. Shoemaker y Eleanor F. Helin, del Instituto de Tecnología de California.

Para un geólogo terrestre, acostumbrado a distinguir una clase de roca de 60 millones de años de antigüedad de otra de 70 millones de años, un error en la edad dentro de un factor de tres apenas si le puede importar mucho más que desconocer en absoluto la edad. Se debe recordar, sin embargo, que el geólogo planetario se enfrenta con el problema de distinguir entre clases de rocas de un millón de años de edad y otras de mil millones de años, es decir, un factor de 1000. Además, como no se cree que haya planetas de antigüedad superior a 4600 millones de años, y como las velocidades de formación de cráteres anteriores a 4000 millones de años eran extremadamente elevadas y disminuyeron rápidamente a lo largo de un corto período de tiempo, las incertidumbres en la datación de una zona por sus cráteres se reducen bastante; las zonas más densamente salpicadas de cráteres tienen probablemente edades comprendidas entre 4000 y 4500 millones de años.

Aparte del sistema Tierra-Luna, no tenemos secuencias estratigráficas fósiles ni edades radiométricas de rocas, sino sólo fotografías de superficies que presentan cráteres. Según ello, sería un progreso impresionante llegar a poder usar los cráteres para correlacionar las historias de los distintos planetas y obtener un método de datación absoluta para la totalidad del sistema solar, aun cuando el método estuviese afectado de errores en un factor de dos o tres. Tal método nos permitiría comparar las velocidades a las que evolucionó cada planeta. Además, podría ser una ayuda para responder a docenas de preguntas fascinantes. Por ejemplo: ¿Se formaron en todos los planetas llanuras de lava semejantes a los mares hace más de tres mil millones de años, como ocurrió en la Luna? O bien: ¿hubo periodos de actividad volcánica dispersos en el tiempo? ¿Estuvo la secuencia de sucesos regida puramente por la masa del planeta y las condiciones térmicas en su interior? ¿Declinó la intensidad de producción de cráteres simultáneamente en todo el sistema solar? ¿Hay alguna indicación de que se produjera un corto impulso de formación catastrófica de cráteres que se extendió a todo el sistema solar interno en algún tiempo particular, como algunos investigadores han supuesto? ¿Están los periodos de erosión, y posiblemente de clima más cálido en Marte, sincronizados con los periodos de clima más cálido en la Tierra? ¿Están sincronizadas las fluctuaciones climáticas en todos los planetas? Si lo están, podrían estar relacionadas con las fluctuaciones en la intensidad de la radiación del Sol, descubrimiento éste que sería de enorme interés para los astrónomos estelares y los biólogos interesados en la evolución.

Las indicaciones que tenemos a mano no son suficientes para responder a esas preguntas en forma definitiva. Las mejores estimas de velocidad de producción de cráteres y los recuentos cuidadosos del número de cráteres, hechos a partir de fotografías, sugieren, sin embargo, la cronología siguiente. Todos los planetas interiores y satélites estuvieron densamente poblados de cráteres hace 4000-4500 millones de años. Los ríos de lava eran corrientes en la Luna y Mercurio hace 3000-4000 millones de años, pero no se han vuelto a registrar desde entonces. En la Luna, los cráteres más jóvenes observables a simple vista desde la Tierra, como el cráter de Tycho con sus brillantes rayos, pueden tener una antigüedad de varios centenares de millones de años. Las regiones de Venus salpicadas de cráteres parecen también mostrar grandes estructuras de impacto que tienen varios miles de millones de años de edad; cráteres más pequeños, de unos 30 km de diámetro o menos, pueden datar de sólo mil millones de años, ya que los más antiguos se habrían erosionado.

Marte presenta un cuadro diferente. Las provincias salpicadas de cráteres parecen albergar cráteres grandes y antiguos, que pueden tener 4000 millones de años de edad, como los de las altiplanicies con cráteres de la Luna y Mercurio. Los cráteres marcianos de menos de 4 km de diámetro tienen vida corta debido a la erosión. La mayor parte de ellos no alcanzan, probablemente, los 3000 millones de años. El hemisferio de Marte menos poblado de cráteres se recubrió, probablemente, de una nueva superficie por lava procedente de una intensa actividad volcánica. Hacia esa misma época, posiblemente hace 1000-2000 millones de años, pudo haber corrido el agua por algunos de los grandes arroyos marcianos. Después, en una fecha tan próxima como 1000 millones de años atrás, o incluso hace sólo unos cientos de millones de años, grandes volcanes, tales como el Monte Olimpo, dieron origen a ríos de lava que no han sido excavados por canales. Tras el periodo de actividad volcánica cesó la erosión más activa, aunque los vientos marcianos continuaron transportando polvo y depositando sedimentos. Algunas de las grandes y más jóvenes formaciones geológicas en Marte son estratos de sedimentos en los polos, que parecen tener una edad de retención de cráteres de apenas 100 millones de años.

Parece que todos los planetas fueron sometidos a intensa producción de cráteres durante los primeros 500 millones de años inmediatos a su formación, y que entonces el volcanismo se convirtió en el proceso dominante de su geología. Los cuerpos menores, tales como la Luna y Mercurio, se enfriaron más de prisa que los otros, y en ellos el volcanismo cesó de actuar. En los planetas interiores mayores, las fuentes internas de calor siguieron siendo importantes. El volcanismo, la actividad tectónica y la erosión atmosférica han destruido parcialmente la primitiva corteza, cubierta de cráteres, de Marte y Venus, y la han destruido casi completamente en la Tierra.

Una cronología definitiva del sistema solar deberá esperar hasta que se tengan muestras de rocas de edad determinable procedentes de muchas provincias de cada uno de los planetas. Tales muestras permitirían a los investigadores determinar la historia del sistema solar con errores de sólo un pequeño porcentaje. Mientras tanto, se puede obtener una cronología provisional del sistema solar buscando fotográficamente asteroides para ampliar nuestro conocimiento del número de cuerpos interplanetarios y, al mismo tiempo, analizando sus órbitas para deducir su historia dinámica. Tales estudios ayudarán a proporcionar mejores estimas de la velocidad de producción de cráteres para todas las partes del sistema solar. Armados con ese conocimiento, podemos llegar a comprender mejor la historia primitiva del sistema solar y el desarrollo no sólo de los planetas interiores y sus satélites, sino también de los gélidos cuerpos del sistema solar exterior.



La productividad de prados mediterráneos

Una aplicación racional de técnicas basadas en el comportamiento de pratenses de clima templado, cultivadas en zona mediterránea, aproxima las producciones reales netas a las potenciales máximas

Arturo Caballero, Francisco Gil y Miguel Berbel

Desde el punto de vista termodinámico, el ecosistema terrestre es un sistema cerrado (sin intercambio) para la materia y abierto para la energía. Ello equivale a decir que los elementos químicos de la biósfera, atmósfera y litósfera no se consumen y cambian solamente de estado energético, absorbiendo energía solar (reacciones endergónicas) para constituirse en moléculas de alto nivel energético y emitiendo energía de naturaleza varia (reacciones exergónicas) al regresar a su estado molecular de contenido energético inferior. Parte de la energía que desprende el sol es desviada por los seres vivos y se utiliza en la realización del trabajo químico o físico necesario para el mantenimiento del proceso vital, tras lo cual vuelve a ser reirradiada a la atmósfera por completo.

El problema de la alimentación para el hombre se reduce, por consiguiente, a canalizar hacia sí el máximo flujo energético solar a través del resto de los componentes de la biósfera.

Los únicos seres capaces de captar directamente la radiación solar son los organismos fotosintetizadores (bacterias fotosintétizadoras, algas azules y vegetales clorofílicos), que reciben por ello el nombre de productores primarios. Gracias a la energía incorporada, estos seres incrementan su biomasa (masa biológica) y sirven de alimento a otros individuos cuyos niveles tróficos son más complejos o superiores. De este modo, la energía solar va pasando, en forma de materia orgánica de alta energía, por todos los eslabones de la cadena alimentaria. En cada una de estas transiciones (productor primario→herbívoro; herbívoro→depredador; depredador→depredador) se pierde, sin embargo, gran parte de la energía inicialmente absorbida por el organismo productor.

Dada esta pérdida de energía en el tránsito de eslabones, se comprende entonces fácilmente que, al menos para los ecosistemas terrestres, la biomasa total de un nivel trófico debe ser mayor que la del nivel que sobre él actúa. En otras palabras, siendo los vegetales superiores cultivados los productores primarios terrestres que están situados en la base de las cadenas alimentarias que culminan en el hombre, una determinada cantidad de estas plantas aportará tanta más energía cuantos menos sean los eslabones intermedios. Esto es, obtenemos más energía de 1 kg de vegetal consumido directamente, que si lo hacemos a través de la carne del animal que lo ha consumido previamente. Por ello, siendo limitada la superficie potencialmente arable de la Tierra (como también lo es la biomasa potencial de productores primarios que el medio acuático puede albergar) se ha de convenir que el número de eslabones de las cadenas tróficas que conducen hasta el hombre limita el tamaño de la población mundial que nuestro ecosistema puede nutrir. De este modo, se llegará indudablemente a la situación límite de tener que alimentarnos exclusivamente de productores primarios enriquecidos con todos los principios que puedan estar sujetos a deficiencia.

No obstante, existen diversos medios por los cuales podrá obtenerse un incremento suficiente de la producción alimentaria total como para permitir aumentos importantes en la población humana sin variaciones sustanciales del tipo de dieta. Para ello es necesaria una potenciación en la producción primaria vegetal que se puede conseguir, entre otras vías, por las siguientes:

a) Selección genética de las variedades cultivadas en orden a la obtención de todos aquellos caracteres hereditarios deseables para una determinada zona y finalidad. Por ejemplo, selección a favor de mayores eficiencias fotosintéticas, menor fotorrespiración, menor consumo de agua y otros nutrientes (fertilizantes), mayor adaptación climática, mayor resistencia a las enfermedades y a los parásitos, enriquecimiento del contenido feculento, graso, proteico o vitamínico, balance entre materia utilizable (frutos, semillas, hoja, raíz) y cosecha total, etc., factores que explicaremos detenidamente más adelante.

b) Reducción racional de la gran diversidad de plantas agrícolas actuales y atención prioritaria, para cada tipo de

LOS CAMPOS EXPERIMENTALES del Departamento de Fisiología Vegetal de la Facultad de Biología de la Universidad de Barcelona están situados en la zona de Pedralbes equidistante, de forma aproximada, de la sierra de Collcerola y del mar. Su suelo presenta una textura francoarenosolimosa, con suficientes propiedades de retención de agua y drenaje, un contenido en carbonato cálcico del 8,5%, nivel de materia orgánica del 7% y pH de 7,7. Se aprecia en esta vista parcial de la instalación diversos cuadrados latinos, de 81 parcelas de 1 m² de superficie cada una, subdivididos en nueve grupos de nueve parcelas que se cortan simultánea o sucesivamente, según lo requieran las necesidades del ensayo, sin que nunca existan en una misma fila o columna dos parcelas sometidas al mismo tratamiento. Los diversos niveles de abonado se aplican semanalmente en fracciones equivalentes y el riego se verifica por un sistema de aspersión homogénea. La recolección de las cosechas se realiza con tijeras, a 3 cm del suelo, y la producción es pesada *in situ* e inmediatamente para obtener su peso fresco; luego, tras desecación en estufa a 110 °C durante 24 horas, se obtiene el peso de la materia seca y el grado de hidratación.

suelo y clima, hacia las variedades que se manifiesten más eficaces en la conversión de la energía solar en biomasa utilizable por el hombre. Junto a ello se debe actuar sobre las cadenas tróficas: cada eslabón que se considere necesario, debe ocuparse por la especie que convierta la energía asimilada de la forma más eficaz.

c) Incorporación mediante material genético externo como episomas, plásmidos o bien mediante mutaciones dirigidas, de caracteres deseables existentes en otras especies, como la capacidad de fijación del nitrógeno elemental, metabolismos de alto rendimiento o vías bioquímicas especiales.

d) Modificación de los parámetros climáticos (microclima), edáficos, nutricionales, de recolección y otros diversos para desarrollar de un modo intensivo las potencialidades endógenas, es decir, las propias de una determinada especie o variedad, basándose para ello en el conocimiento lo más profundo posible de la fisiología característica de la misma.

El ámbito de este artículo queda contenido por completo dentro de este último tipo de actuación: plantas de cosecha continua que han sido sometidas a regímenes especiales de riego, abonado y períodos de recolección o corte.

Es innegable que, en el panorama actual agrícola, las plantas pratenses poseen un alto valor económico como vehículo de tránsito de la energía solar al ganado doméstico de alto rendimiento en la conversión de la misma en alimento para el hombre. De ahí se desprende el sumo interés de la potenciación al máximo de la producción de los prados.

En su contexto geobotánico, la palabra prado se utiliza para denominar el tipo de vegetación natural constituido por gramíneas (avena, grama, espigueta, cañuela, raigrás, etc.), papilionáceas (trébol, alfalfa, etc.), y otras plantas herbáceas de carácter mesofítico, es decir, adaptadas a situaciones intermedias de humedad. Las condiciones edáficas de las zonas de prado, así como la propia competencia de sus especies características, impiden la germinación de las semillas o el desarrollo de otros vegetales de mayor porte. Sin embargo, en sentido más amplio, se puede aplicar el mismo término para designar una formación artificial, de características fisionómicas similares, de flora espontánea o cultivada. Los resultados que se detallan en este artículo se han obtenido, precisamente, en prados artificiales cultivados, de siega continua, situados en los Campos Experimentales del Departamento de



PLANTA C_3 (*Festuca elatior* o festuca) y planta C_4 (*Cynodon dactylon* o grama). Las diferencias morfológicas entre estas dos especies no son demasiado acentuadas ya que ambas pertenecen a la familia gramíneas. Sin embargo, existen claras diferencias fisiológicas: sus metabolismos de incorporación y reducción del CO_2 son esencialmente distintos. También son patentes ciertas diferencias anatómicas, la más importante de las cuales reside en que el mesófilo de la planta C_3 está dividido en un parénquima en empalizada y en un parénquima lagunar, mientras que la C_4 posee el mesófilo dispuesto en corona alrededor de la vaina del haz conductor. (Redibujado de Buendía, 1965.)

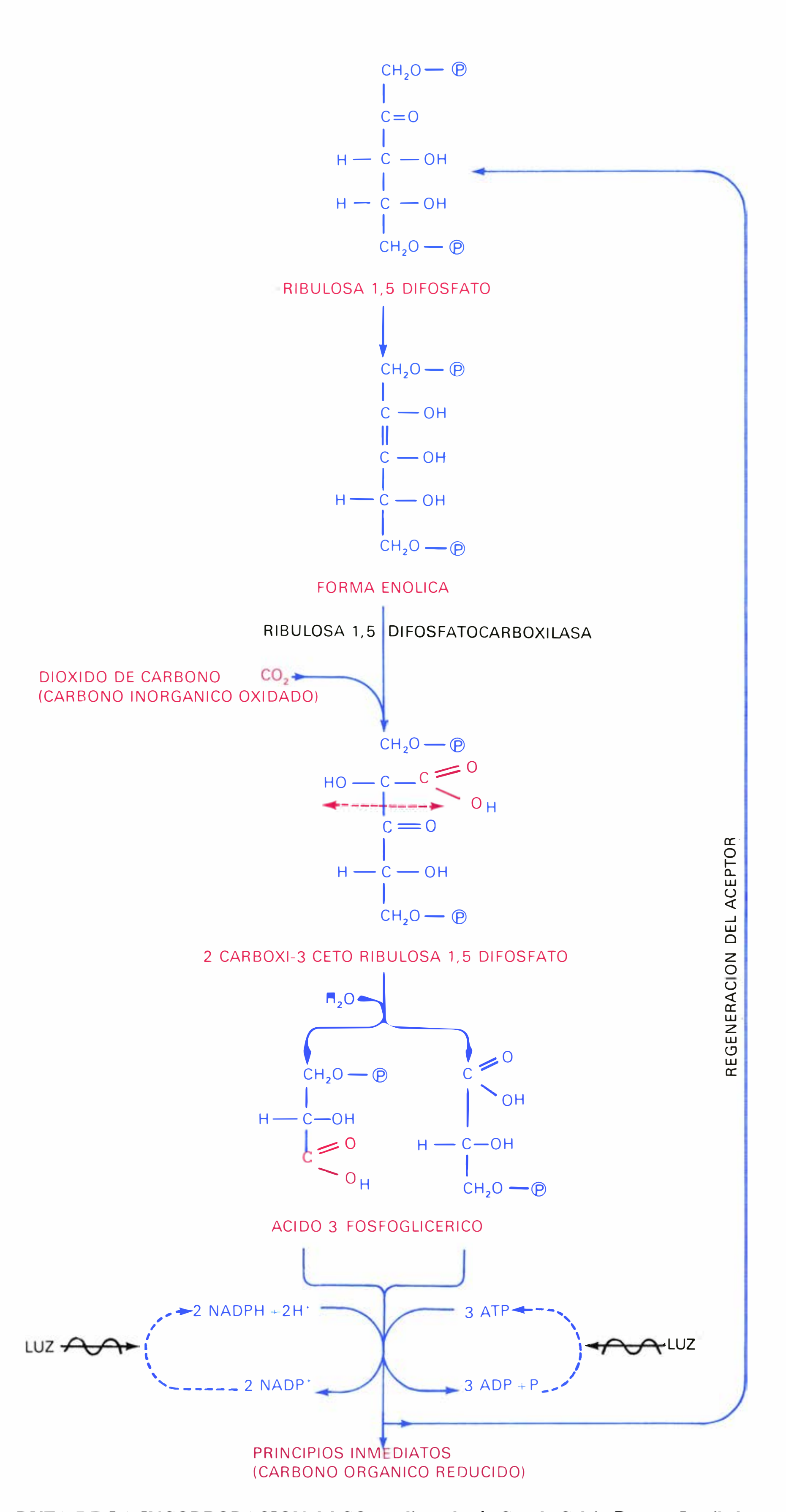
Fisiología Vegetal de la Facultad de Biología de la Universidad de Barcelona.

Las variaciones climáticas de la localidad de experimentación se pueden considerar típicas de la zona climática mediterránea, en la que se halla localizada geográficamente. Los parámetros del diagrama climático correspondiente arrojan valores de 16° C para la temperatura media anual y de 600 mm para la precipitación media en el mismo lapso de tiempo, con mínimas extremas observadas (mediciones efectuadas durante 20 años) de –4.3° C. Existen seis meses susceptibles de tener mínimas absolutas inferiores a los 0° C (de noviembre a abril) y los máximos de pluviosidad media se sitúan en mayo (unos 60 mm) y noviembre (alrededor de 95 mm). La época de temperatura media más alta se localiza en agosto (cerca de 22° C) y la de menor temperatura en enero (9° C aproximadamente). De todo ello se desprende que existe, en cada año natural, un período relativamente árido, localizado en los meses de junio, julio y agosto, que separa a dos períodos relativamente húmedos, uno situado en el intervalo invierno-primavera y otro otoñoinvernal. Las variaciones estacionales son, pues, inviernos fríos, primaveras y otoños suaves, con pluviosidad suficiente, y veranos cálidos, luminosos y secos.

Las condiciones edáficas de la citada instalación experimental se situaron en una textura francoarenosolimosa, con suficientes propiedades de retención de agua y drenaje, un contenido en carbonato cálcico del 8.5 %, nivel de materia orgánica del 7 % y pH de 7,7.

Con estas características de suelo y clima y con el material cespitoso a que antes hemos aludido, los esfuerzos de nuestro equipo científico y técnico se han dirigido durante los últimos años hacia dos finalidades muy concretas: 1) la obtención y mantenimiento de producciones totales que alcancen los máximos teóricos calculados y 2) el logro de una mayor uniformidad de distribución de la biomasa recolectada respecto a las variaciones estacionales.

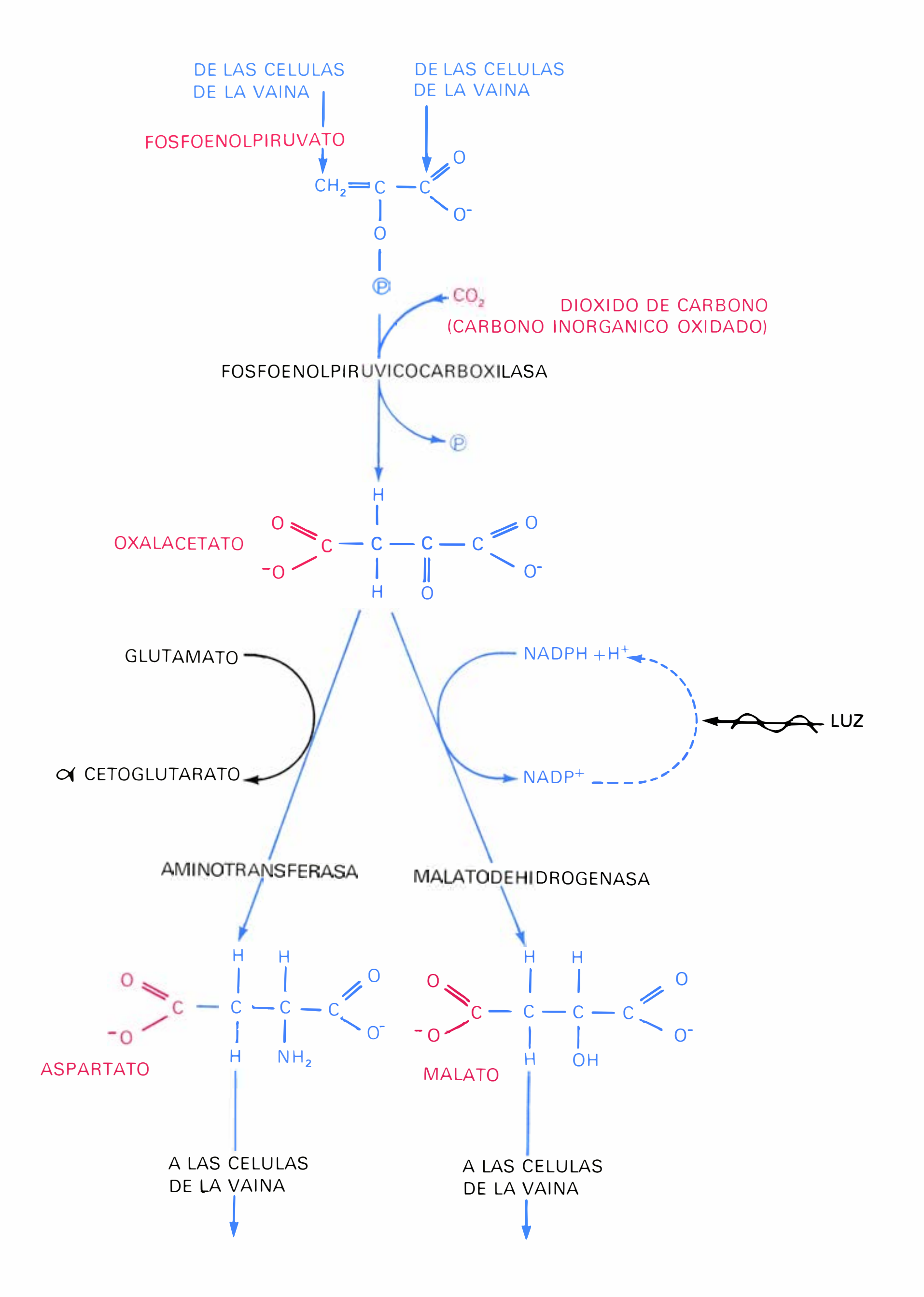
La distribución, a lo largo del año, del crecimiento de las plantas pratenses, presenta un carácter estacional muy marcado. Ello se debe a las distintas características constitucionales o genéticas de cada especie o variedad, en relación con los factores climáticos que varían con la sucesión de las estaciones. Aunque, en principio, la respuesta de desarrollo es bastante específica, existe, sin embargo, multitud de rasgos comunes que posibilitan el establecimiento



RUTA DE LA INCORPORACION del CO_2 mediante la vía C_3 o de Calvin-Benson. La ribulosa, 1,5 difosfato capta una molécula de dióxido de carbono resultando de esta carboxilación una molécula hexacarbonada que se dismuta en dos de ácido 3 fosfoglicérico. Esta reacción química está catalizada por un enzima muy especial, la ribulosa 1,5 difosfatocarboxilasa, que es capaz de actuar tanto sobre el CO_2 (carboxilación) como el O_2 (oxigenación). El ácido 3 fosfoglicérico sufre una serie de transformaciones en las cuales su grupo carboxílico se reduce por la acción del NADPH + H^+ y del ATP, originando una triosa (azúcar de tres átomos de carbono) que da lugar a los principios inmediatos, así como a la molécula aceptora que inicia de nuevo el ciclo. El gasto metabólico total corresponde a 2 NADPH + 2 H^+ y 3 ATP por cada molécula de CO_2 incorporada.

de dos grandes grupos en las principales plantas cultivadas. Uno, el de las plantas C_3, que integra a la mayoría de las especies, se caracteriza por presentar óptimos térmicos de producción próximos a los 20° C, requerir intensidades luminosas menores que la solar para la asimilación fotosintética óptima y necesitar suministros hídricos elevados. Otro, el de las plantas C_4, mucho más reducido en amplitud, que precisa óptimos térmicos más elevados (entre 30° y 35° C), luminosidades máximas (incluso superiores a la solar) y que manifiesta menores exigencias de agua.

Las diferencias esenciales entre el comportamiento frente a los factores ambientales de las plantas C_3 y C_4 sólo pueden comprenderse íntimamente cuando se conocen los mecanismos bioquímicos que se sitúan en la base de fenómeno fotosintético. Para ello, consideramos imprescindible una breve descripción de las rutas de reducción (incorporación de electrones) y asimilación del anhídrido carbónico en los vegetales.



LA VIA DE LAS PLANTAS C_4 mantiene una separación física entre dos carboxilaciones; la primera de ellas tiene lugar en las células del mesófilo en corona y se verifica mediante el CO_2 ambiental. La molécula aceptora (fosfoenolpiruvato) incorpora el dióxido de carbono en una reacción catalizada por la fosfoenolpiruvatocarboxilasa originándose oxalacetato. Este ácido dicarboxílico tetracarbonado puede entonces sufrir dos trasformaciones: *a)* reducción asociada al $NADPH + H^+$ o al $NADH + H^+$ regida por la malato dehidrogenasa o *b)* transaminación a aspartato condicionada por una aminotransferasa asociada al sistema glutamato / α cetoglutarato. En cualquiera de los dos casos, el producto formado (malato o aspartato) posee en un grupo carboxílico el CO_2 capturado y se acumula en las células de la corona constituyéndose un gradiente de concentración a favor del cual estos ácidos se difunden a las células de la vaina foliar de los haces conductores. De modo parecido, las plantas crasas (metabolismo CAM) mantienen una separación entre la primera carboxilación, nocturna y con los estomas abiertos, y la segunda, diurna y por decarboxilación de ácidos cuando los estomas están cerrados para evitar la excesiva transpiración.

De un modo muy esquemático, se puede definir la fotosíntesis como el proceso de conversión de la energía luminosa en energía química, la cual se almacena en moléculas orgánicas de alto nivel energético. Parte de la luz se acumula (52 Kcal/mol) como moléculas de alto poder reductor, a saber nicotinamin adenindinucleótidofosfato ó reducido o $NADPH + H^+$, y parte (7 Kcal/mol) como moléculas de alta energía, es decir, adenosintrifosfato o ATP, que suministran la diferencia de potencial necesaria para la reducción del CO_2. La incorporación y posterior reducción del CO_2 a nivel de materia orgánica (CH_2O) almacena unas 112 Kcal/mol y requiere del concurso de 4 electrones que son suministrados por dos moléculas de $NADPH + H^+$ y de 3 moléculas de ATP.

Las plantas C_3 verifican una serie de reacciones de captura y reducción del dióxido de carbono que reciben el nombre de vía de Calvin-Benson o vía C_3 (porque el primer compuesto estable de la carboxilación es una molécula de tres átomos de carbono). En este proceso, la molécula aceptora del anhídrido carbónico es un azúcar de tipo cetónico de cinco átomos de carbono, la ribulosa 1,5 difosfato o 1,5 RuDP, y el sistema enzimático principal es la ribulosa 1,5 difosfatocarboxilasa que, simultáneamente, cataliza la carboxilación y la dismutación de la molécula resultante de seis átomos de carbono (2 carboxi-3ceto ribulosa 1,5 difosfato) en dos moléculas idénticas, de tres átomos de carbono, es decir, dos moléculas de ácido 3 fosfoglicérico (3 PGA). Una de estas dos moléculas formadas posee, en su grupo carboxílico, el CO_2 capturado.

El 3 PGA, así formado, sufre entonces una reducción mediante $NADPH + H^+$ y ATP, procedentes de la fase luminosa de la fotosíntesis, y se convierte en aldehido 3 fosfoglicérico (3 PGA1), que se integra finalmente en las vías metabólicas de síntesis de principios inmediatos y, también, en el ciclo de regeneración del aceptor (1,5 RuDP).

Las plantas C_4 realizan un proceso de captación y reducción del CO_2 que ofrece variantes esenciales respecto a los vegetales C_3. En realidad, en estas plan-

tas coexisten un proceso específico, propiamente C_4, porque la primera molécula estable es un ácido dicarboxílico tetracarbonado (4 carbonos), y el proceso general que ya se ha especificado para las plantas C_3, es decir, la carboxilación de la 1,5 RuDP. Pero, además, existen también diferencias estructurales entre ambos grupos fisiológicos de vegetales. En efecto, el mesófilo o parte interna de las hojas de las plantas C_3 está compuesto por células más o menos prismáticas y alargadas, dispuestas en una o más capas densas, con escasos espacios intercelulares, el parénquima en empalizada, y células más isodiamétricas, irregulares, dispuestas con amplios espacios entre sí para facilitar la circulación del agua y los gases, el parénquima lagunar. En cambio, en las plantas C_4 existe una particular anatomía ("kranz" o en corona) que se revela por una disposición radial del parénquima clorofílico alrededor de la vaina de los haces conductores de las hojas. Dicha vaina está constituida por células que, a diferencia de lo habitual en la vaina de las plantas C_3, presentan cloroplastos abundantes.

Mientras que los cloroplastos de las células del mesófilo de las plantas C_3 son totalmente típicos, en las plantas C_4 existen dos tipos de cloroplastos.

a) Cloroplastos de la corona. Muy esquemáticamente se puede decir que son parecidos a los cloroplastos de las plantas C_3, pero carecen de 1,5 ribulosadifosfatocarboxilasa y, por consiguiente, no son capaces de sintetizar ni acumular almidón. Poseen, sin embargo, altos niveles del enzima fosfoenolpirúvicocarboxilasa.

b) Cloroplastos de la vaina. Carecen casi por completo de grana y, por lo tanto, de fotosistema II, por lo que son incapaces de acumular potencial reductor ligado a la luz. Poseen 1,5 ribulosadifosfatocarboxilasa y son capaces de sintetizar y acumular almidón.

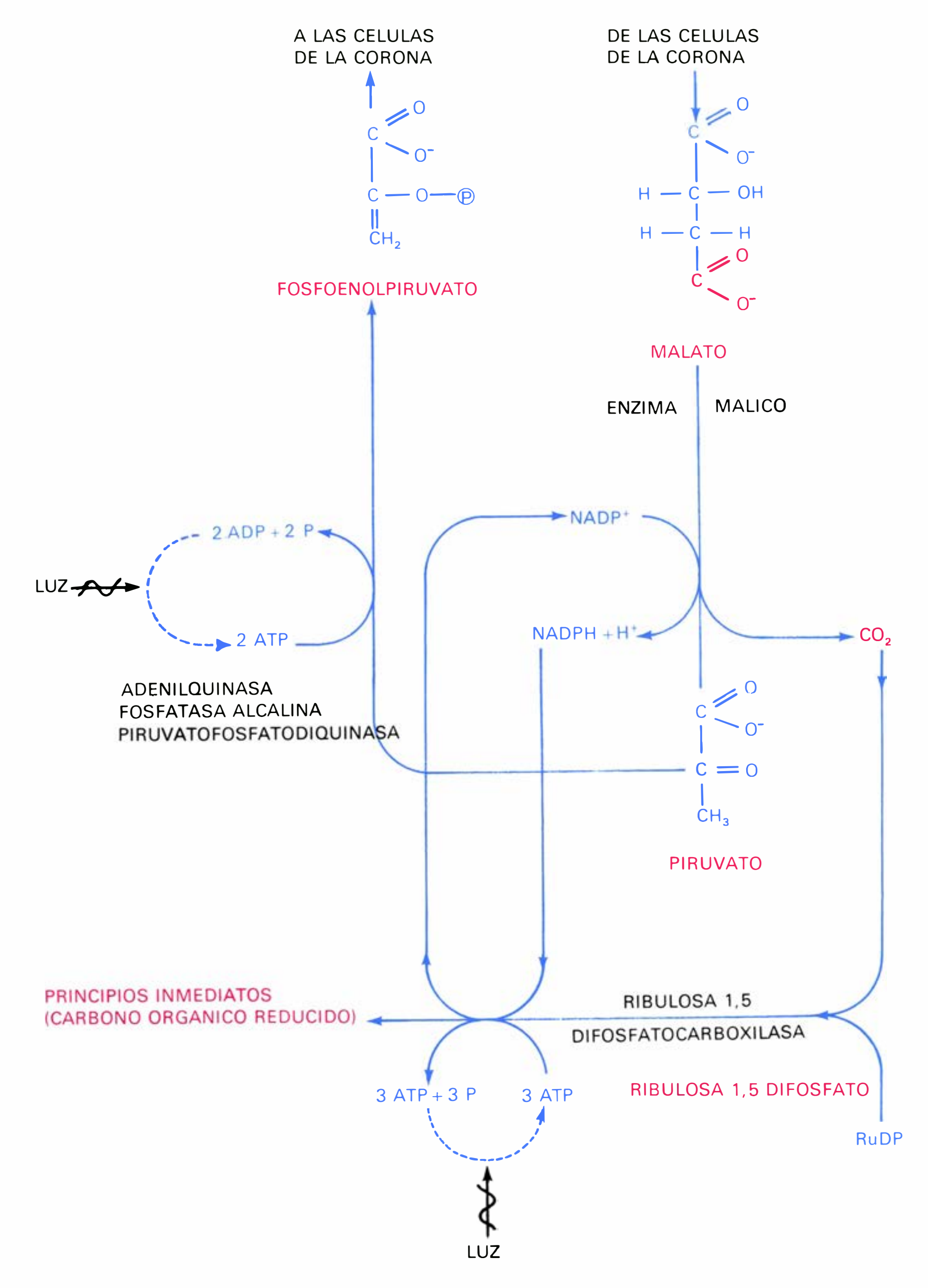
En la ruta C_4, la molécula aceptora, fosfoenolpiruvato o PEP, se carboxila en las células de la corona mediante la fosfoenolpirúvicocarboxilasa produciéndose oxalacetato, el cual puede seguir dos vías:

α) Reducción a malato, con el consiguiente gasto de poder reductor en forma de $NADPH + H^+$ o de nicotaminadenindinucleótido ($NADH + H^+$) procedentes, directa o indirectamente, de la fase luminosa de la fotosíntesis.

β) Transaminación a aspartato, reacción catalizada por una aminotransferasa o transaminasa.

La acumulación de los productos anteriores (malato o aspartato) en las células de la corona origina un gradiente de concentración a favor del cual estos ácidos dicarboxílicos penetran en la vaina, vía plasmodesmos.

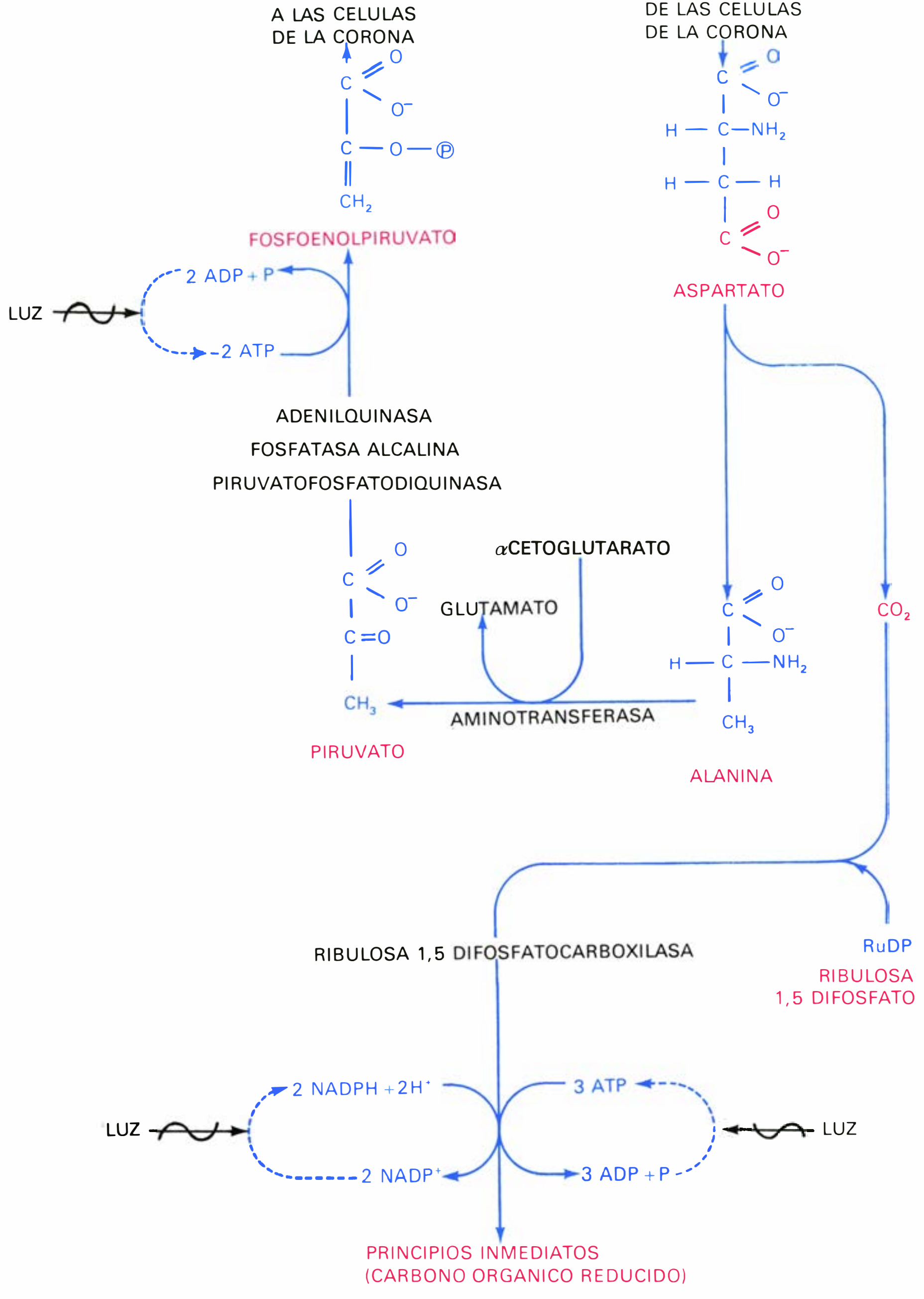
En las células de la vaina se procede a la decarboxilación de estos compuestos. En el caso del malato, la reacción se cataliza por el enzima málico que libera CO_2, piruvato y potencial reductor ($NADPH + H^+$ o $NADH + H^+$). El CO_2 desprendido, gracias al potencial reductor liberado y con el concurso de adenosintrifosfato (ATP) procedente de la fotofosforilación de los cloroplastos de las células de la vaina, sigue entonces



EN LAS CELULAS DE LA VAINA, el malato procedente de la corona sufre la actuación del enzima málico y se descompone en CO_2 y piruvato, en una reacción en la que se acumula poder reductor ($NADPH + H^+$). En realidad, el proceso se puede considerar como una inyección de dióxido de carbono y un transporte aparente de poder reductor desde la corona a la vaina; es un fenómeno importante por dos razones, ya que al aumentar la concentración de CO_2 en la vaina se mejora el rendimiento de la ribulosa 1,5 difosfatocarboxilasa y, por otro lado, es imprescindible el aporte de equivalentes de reducción porque los cloroplastos de las células de la vaina carecen de grana y, por consiguiente, no pueden acumular $NADPH + H^+$ ligado a la luz. El CO_2, el $NADPH + H^+$ y el ATP originado en las reacciones fotoquímicas de las células de la vaina se integran entonces en el ciclo de Calvin-Benson o ruta C_3 donde se da la segunda carboxilación. Por consiguiente, las plantas C_4 realizan, en realidad, una secuencia de las rutas C_4 y C_3. El piruvato originado, con gasto de 2 ATP y el concurso de la piruvatofosfatodiquinasa, la fosfatasa alcalina y adenilquinasa se convierte en fosfoenolpiruvato que, en las células de la corona, actúa de nuevo de molécula aceptora.

la vía C_3 o ciclo de Calvin-Benson, siendo dicho proceso típico de las células de la vaina que, así, se asemeja a los tejidos fotosintéticos de las plantas C_3.

El piruvato residual, con gasto de la energía de 2 ATP y la acción de la fosfatopiruvatodiquinasa, la pirofosfatasa alcalina y la adenilquinasa se transforma de nuevo en PEP que, saliendo de la vaina y acumulándose en la corona, puede iniciar de nuevo el ciclo. En el caso del aspartato se procede en la vaina a una decarboxilación que desprende CO_2 y el aminoácido alanina. El dióxido de carbono, con gasto de ATP y de NADPH + H^+ (los cloroplastos de la vaina de las especies que acumulan aspartato tienen siempre algún granum o zona de síntesis de potencial reductor) sigue la vía C_3, como en el caso anterior. La alanina, por transaminación, se convierte en piruvato que, a su vez, se transforma, por el procedimiento antes indicado, en PEP aceptor.



LAS PLANTAS C_4 que siguen la vía del aspartato ofrecen una interesante variación respecto a las que verifican el metabolismo del malato. El aspartato procedente de las células de la corona se convierte en alanina por decarboxilación, liberándose CO_2. En la vaina, este dióxido de carbono se integra en la vía C_3 mediante la ribulosa 1,5 difosfatocarboxilasa y es reducido por la acción del NADPH + H^+ y del ATP. Ambos, el potencial reductor y el potencial energético auxiliar, se logran asociados a las reacciones fotoquímicas de los cloroplastos de las células de la vaina, ya que éstos, aunque en poco número, poseen algún granum. La alanina, por una aminotransferasa y el sistema α cetoglutarato / glutamato, se convierte en piruvato que, como en la vía del malato, se transforma, con el gasto de dos ATP, en la molécula aceptora (fosfoenolpiruvato). Por consiguiente y comprendiendo el balance energético total, en las plantas C_4 (vía malato o vía aspartato) es de 2 NADPH + 2 H^+ y 5 ATP por CO_2 reducido a materia orgánica (a nivel de CH_2O) mientras que en la vía C_3 el consumo sólo es de 2 NADPH + 2 H^+ y 3 ATP. A pesar de que la ruta C_4 es energéticamente menos eficaz, se revela como más productora, a temperaturas altas y luminosidad intensa, ya que mejora extraordinariamente la eficacia de la ribulosa 1,5 difosfatocarboxilasa y, ligado a ello, reduce mayoritariamente la fotorrespiración determinando un incremento de fotosíntesis neta.

En la ruta C_4, por consiguiente, se localizan en el espacio dos tipos de procesos: una carboxilación muy eficaz, en la corona, y una decarboxilación, nueva carboxilación y reducción del CO_2 liberado, en la vaina. El primero de estos dos fenómenos está adaptado a altas intensidades luminosas y a temperaturas relativamente altas.

El consumo energético del ciclo de Calvin-Benson es de 3 ATP por cada CO_2 reducido. En el ciclo C_4, a este consumo, hay que añadir 2 ATP para la regeneración del PEP, de lo que resulta un gasto de 5 ATP por CO_2 asimilado. Dicho de otro modo, la ruta C_4 es energéticamente menos eficaz que la ruta C_3. Sin embargo, el flujo de CO_2 a la vaina, procedente de la decarboxilación de aspartato o malato, da una concentración de CO_2 de 1.5 a 2 veces superior a la que tiene el dióxido de carbono atmosférico en el mesófilo de las plantas C_3. Ello determina que aumente extraordinariamente el rendimiento de la 1,5 ribulosadifosfatocarboxilasa (enzima que tiene muy poca afinidad por el sustrato) por lo que, a pesar de todo, las plantas C_4 tienen una mayor productividad (producción respecto a la biomasa) que las plantas C_3. Dicho efecto es comparable al aumento de producción de las plantas C_3 cuando son abonadas con CO_2.

La producción vegetal viene condicionada por el balance entre fotosíntesis bruta y respiración que, en cierto modo, son dos fenómenos inversos. El consumo de substancias oxidables, como los azúcares, en los procesos respiratorios, merma el acervo de asimilados fotosintéticos. En condiciones de buena iluminación, temperaturas adecuadas y regular suministro de nutrientes, la componente fotosintética supera al gasto respiratorio, estableciéndose el concepto de fotosíntesis neta como la suma algebraica de los dos procesos considerados.

La determinación de la intensidad respiratoria en condiciones de oscuridad no ofrece dificultad alguna, dado que, en estas condiciones, no existe fotosíntesis. Sin embargo, bajo iluminación, sólo se puede medir directamente la fotosín-

tesis neta, con lo cual surge el problema de dilucidar si el proceso respiratorio prosigue a la luz con la misma o distinta intensidad que en la oscuridad.

En general, los vegetales presentan, a la luz, una mayor intensidad respiratoria, que se puede demostrar por muy diversos métodos y que da origen al concepto de fotorrespiración. Para ciertos autores, la fotorrespiración corresponde al incremento respiratorio producido por la luz y, para otros, a la totalidad del proceso respiratorio en condiciones de iluminación. En sentido restringido, se puede definir la fotorrespiración como el incremento de intensidad respiratoria, desencadenado por la luz en los órganos fotosintéticos, verificado sobre sustratos producidos inmediatamente por la actividad fotosintética y detectable por el desprendimiento de CO_2.

La fotorrespiración se acentúa con el aumento de temperatura y de la concentración de oxígeno, mientras que se atenúa, con la consiguiente ganancia de fotosíntesis neta, con el incremento de la concentración de anhídrido carbónico. Por todo ello, para cada temperatura fisiológica y en condiciones no limitantes de humedad e iluminación, existe una concentración de anhídrido carbónico (punto de compensación para el CO_2) que iguala la magnitud del proceso fotosintético con la de los fenómenos respiratorios. En condiciones de punto de compensación la fotosíntesis neta es, pues, nula.

Las plantas pertenecientes al grupo C_3 presentan una fotorrespiración elevada que condiciona un alto punto de compensación (cerca de 50 ppm de CO_2 a unos 25° C). En el caso de las plantas C_4, se ha demostrado que la componente fotorrespiratoria o no existe, o es de una intensidad mínima. Todo ello conduce a puntos de compensación muy bajos (0-5 ppm). Si estos vegetales no presentan desprendimiento postiluminatorio y, sin embargo, sí poseen el acervo enzimático propio de la fotorrespiración, se deberá admitir que el CO_2 desprendido por fotorrespiración es reutilizado de nuevo, antes de salir del mesófilo foliar, dada la gran capacidad de absorción del enzima fosforenolpirúvicocarboxilasa de las células de la corona. Otra explicación verosímil, y seguramente acoplada con la anterior, es que, al ser más elevada la concentración de CO_2 en las células de la vaina de las plantas C_4, se disminuye la magnitud de la fotorrespiración por el propio efecto Warburg.

En condiciones naturales, la máxima producción de las plantas C_3 resulta ser del orden de 28-30 mg CO_2 $dm^{-2}h^{-1}$ (miligramos de CO_2 absorbido por decímetro cuadrado de hoja y hora) para vegetales cultivados de gran rendimiento (cereales como el trigo, el arroz, la avena, plantas hortícolas como el tomate, las patatas, etc). Los máximos de producción en las plantas C_4 oscilan entre 60-80 mg CO_2 $dm^{-2}h^{-1}$ (caña de azúcar, maíz, sorgo, etc.).

Las diferencias pueden interpretarse en razón de los siguientes criterios: las plantas C_4 no presentan fotorrespiración y, en ellas, el ciclo de Calvin es más eficaz debido a la mayor concentración de CO_2 que aumenta el rendimiento de la 1,5 ribulosadifosfatocarboxilasa, y, por consiguiente, la incorporación del dióxido de carbono en substancias asimiladas.

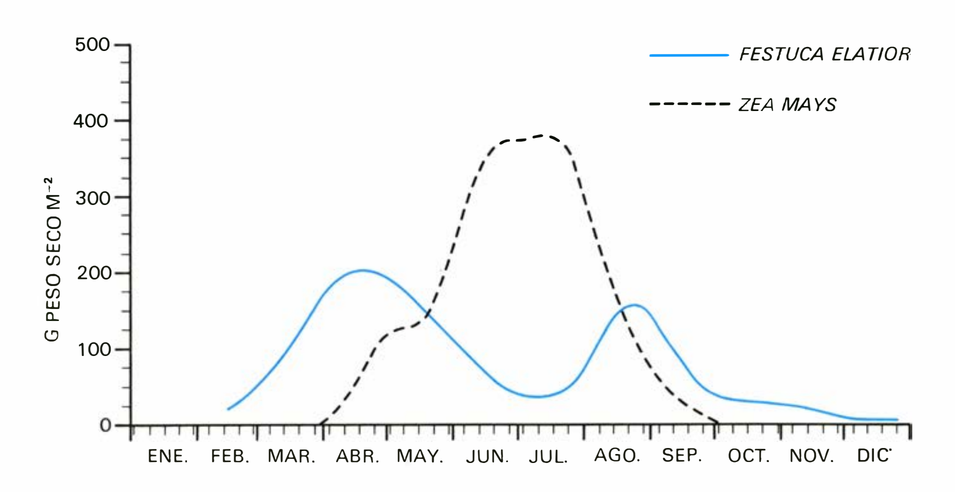
Por todo ello, las plantas C_4 representan una adaptación genética a climas de alta intensidad luminosa, temperaturas relativamente altas (30°-35° C) y con una no muy grande disponibilidad hídrica. No obstante, la colonización de las zonas áridas no se ha llevado a cabo exclusivamente por plantas C_4, sino que se ha verificado por vegetales de ambos grupos bioquímicos. Tampoco podemos considerar que el atributo C_3 o C_4 haya tenido un valor genético selectivo sobre la especiación vegetal, dado que, en un mismo género, y en especies afines, se encuentran representantes de los dos grupos.

La productividad puede definirse como el incremento de peso en materia seca por unidad de superficie cultivada; normalmente, se expresa en kilogramos por hectárea. La producción neta es el resultado de la integración de diversos factores endógenos (genéticos, fisiológicos, anatómicos, etc.) con los ambientales (suelo-clima) y con los correspondientes a las prácticas del cultivo.

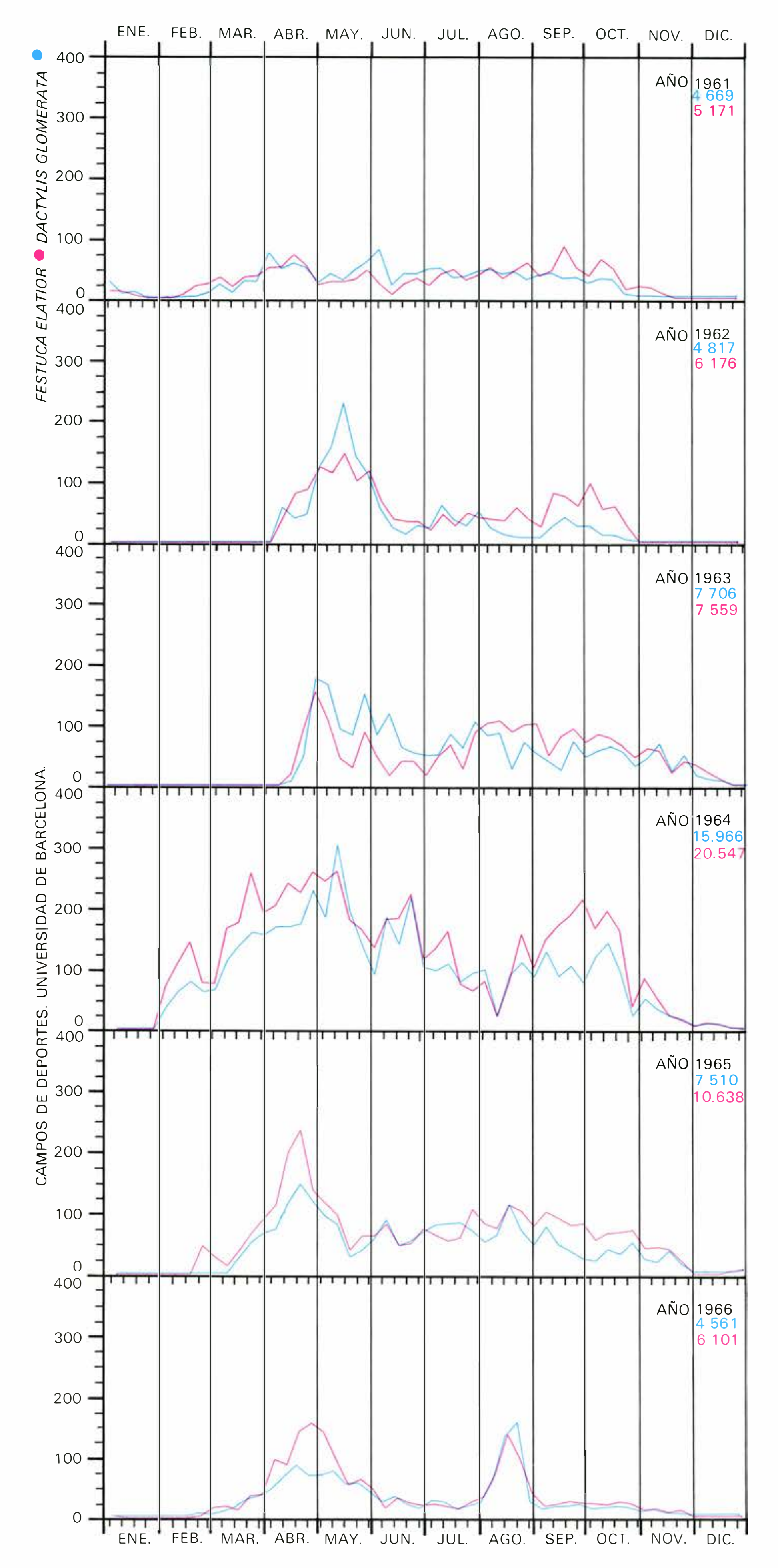
La asimilación fotosintética del anhídrido carbónico constituye el proceso principal en la producción de materia orgánica por las plantas. Pero la fotosíntesis exhibe variaciones considerables, entre otros aspectos, porque las comunidades vegetales constituyen sistemas dinámicos que ofrecen múltiples cambios estacionales y de ciclos biológicos. La capacidad de la planta para adaptarse a nuevas condiciones, la rapidez para desarrollar su cobertura foliar y las características de la redistribución de los asimilados, desde los órganos productores a los consumidores, son ejemplos de aspectos que repercuten decisivamente en la capacidad fotosintética. Así, el curso del crecimiento y del desarrollo afecta al comportamiento fotosintético y, por lo tanto, a la producción.

En los ambientes (suelo-clima) cuyas capacidades para suministrar nutrientes y agua sean limitadas, la producción puede resultar muy mermada. Pero incluso cuando estos recursos no pueden considerarse limitantes originan diversas repercusiones en la producción.

La productividad de las plantas del grupo C_3 se considera, en general, más baja que la productividad de las pertenecientes al C_4, siendo sus tasas de producción (peso seco m^{-2} semana^{-1}) la mitad o, incluso, la tercera parte. Sin embar-



LAS VARIACIONES DE TEMPERATURA Y HUMEDAD características del clima mediterráneo determinan que, en general, las especies pertenecientes al grupo fisiológico C_3 encuentren condiciones favorables para su crecimiento en primavera y otoño; por ello, la distribución anual de su producción presenta un caracter bimodal, con dos máximos separados por una depresión estival. En cambio, las plantas del grupo C_4, solamente hallan un ambiente adecuado en verano, de modo que su perfil presenta un único máximo centrado en el estío y su periodo vegetativo es, en nuestra latitud, más corto. La producción de ambos tipos de vegetales queda complementada a lo largo del tiempo. En la gráfica, producción en peso seco (g m^{-2}) respecto al tiempo en una planta C_3 (*Festuca elatior*, en azul) y otra C_4 (el maíz, *Zea mays*, en negro) de los Campos Experimentales.



go, autores como Gifford (1974) presentan ejemplos en los que no pueden encontrarse estas diferencias a favor de los vegetales C_4, sobre todo, si se considera cada especie en su ambiente apropiado. Las producciones anuales pueden quedar condicionadas porque los factores ambientales favorezcan durante más tiempo a una especie que a otra.

Así, por ejemplo, las variaciones de temperatura y humedad características del clima mediterráneo determinan que, en general, las especies pertenecientes al grupo bioquímico C_3 encuentren solamente un ambiente adecuado en primavera, principios de verano y otoño. Por ello, la distribución anual de su crecimiento presenta un carácter netamente bimodal, con dos máximos de producción separados por la depresión estival, debida al calor y a la sequedad, y la inhibición invernal, condicionada por las bajas temperaturas. Por el contrario, las especies del grupo fisiológico C_4 hallan solamente condiciones ambientales convenientes en la época estival, de modo que su perfil ofrece un único máximo principal centrado en el verano, con un período vegetativo muy corto. De ello resulta que ambos tipos de especies imbrican sus producciones dentro del año y, si bien estos cultivos mixtos y los rotatorios ofrecen en ocasiones interés económico o aplicado considerable, ocasionan dificultades en cuanto a su manejo y, sobre todo, ofuscan el estudio de las capacidades individuales de producciones reales y potenciales.

Las pratenses más interesantes para el cultivo de climas templados se encuentran entre las representantes del grupo C_3, ya que las plantas C_4, con máximos de producción estival de mayor rendimiento por superficie cultivada, tienen, en nuestra zona, una duración tan escasa

PERFILES DE PRODUCCION de dos pratenses cultivadas en idénticas condiciones de campo. Se observan producciones crecientes desde 1961 a 1964 y decrecientes de 1964 a 1966. La primera oscilación es debida al efecto del abonado (1961, sin abono, 1962, 100 kg N ha^{-1} $año^{-1}$; 1963, 190 kg N ha^{-1} $año^{-1}$; a partir de 1964, 250 kg N ha^{-1} $año^{-1}$). La pérdida de rendimiento posterior fue causada por degradación del campo. Es también apreciable una alternancia anual de perfiles bimodales y multimodales y, además, queda reflejada la mayor similitud entre las distribuciones de producción de las dos especies distintas en un mismo año, que en la misma especie en los años sucesivos, con lo que se demuestra, por consiguiente, el gran efecto de los factores ambientales (suelo-clima) en el crecimiento cuando se aplican programas habituales de cultivo. En las gráficas se representan en rojo los valores correspondientes a *Festuca elatior* y en azul los pertenecientes a *Dactylis glomerata*. Para cada año se expresa la producción de materia seca obtenida por siega en kilogramos ha^{-1} $año^{-1}$.

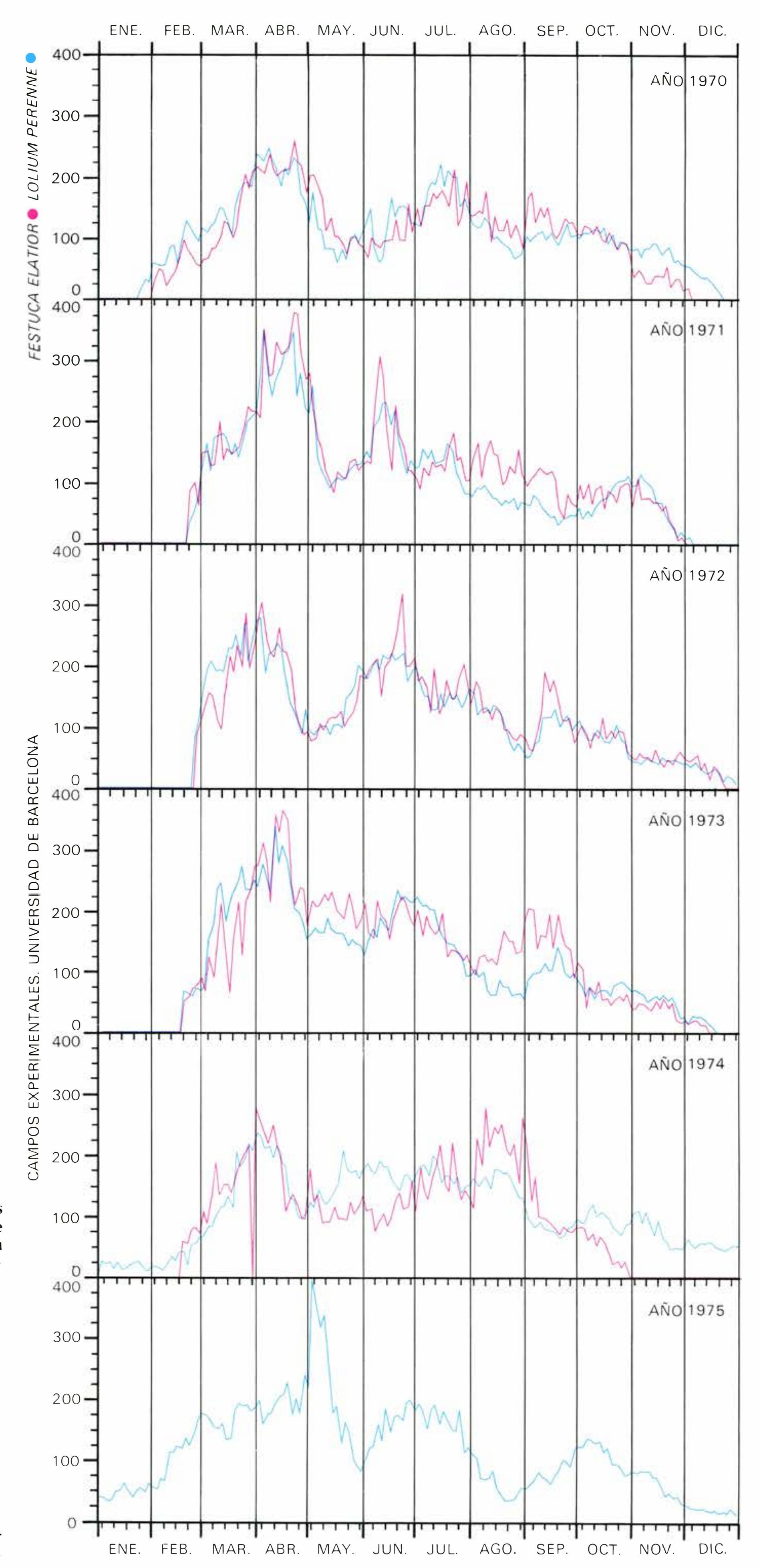
del período vegetativo que su producción anual resulta bastante inferior.

Loomis y Hunt (1976) asignan gran importancia al grado de disponibilidad de los nutrientes en la integración del crecimiento y de la producción fotosintética. En muchos ambientes agrícolas, los elementos nutritivos asimilables y las disponibilidades de agua deciden el crecimiento y la cosecha. Los sistemas cultivados, si se comparan con los naturales, no suelen ser precisamente pobres en recursos, debido a que las prácticas de abonado, riegos, labores, etc. contribuyen a crear ambientes ricos. A pesar de ello, la utilización de especies o variedades de elevada capacidad productiva exige niveles de abonado muy superiores y más controlados que los característicos de los sistemas naturales.

Entre todos los nutrientes, el nitrógeno es, sin ninguna duda, el más decisivo, si bien, para conseguir una nutrición equilibrada, es necesario acompasar simultáneamente los niveles de otros elementos (fósforo, potasio, etc.)

Loomis y Gerakis (1975) indican que especies agrícolas más eficaces del grupo C_4, en las cuales el nitrógeno supone un 2 % de su peso seco, pueden precisar cantidades del orden de 1600 kg de N por hectárea y por año, para poder llegar a producir los máximos de biomasa de 80 Tm de peso seco por hectárea y por año, y que las especies más productoras de los grupos C_3 y C_4, cultivadas en clima templado, exigen unos 800 kg N para alcanzar las 40 Tm de peso seco por hectárea y año.

Las plantas que hemos ensayado, con excepción de las fijadoras de nitrógeno (leguminosas), es decir, capaces de asimilar el nitrógeno atmosférico, sólo pueden proporcionar elevados rendimientos fotosintéticos bajo aplicaciones masivas de abono nitrogenado. La mayor parte de los cultivos agrícolas habituales se



PRODUCCIONES ESTACIONALES de dos gramíneas sometidas a régimen intensivo de riego (unos 20 litros por metro cuadrado cada dos días como máximo) y altas tasas de abonado nitrogenado (640-2000 kg N ha^{-1} $año^{-1}$). Estas producciones, referidas solamente al material cosechado por siega, son bastante más altas que las conseguidas en cultivos anteriores. Se aprecia claramente la gran similitud de perfil entre las curvas de ambas especies para cada año determinado, aunque con un caracter estacional mucho más atenuado que en los cultivos de 1961 a 1966. A diferencia de la gráfica anterior, estos campos no muestran ningún síntoma de agotamiento. Las dos series de datos tienen desigual duración ya que las experiencias con *Festuca elatior* se dieron por concluidas en 1974, mientras que las de *Lolium perenne* prosiguen aun en la actualidad. En rojo produción de *Festuca elatior;* en azul, de *Lolium perenne.*

LOLIUM PERENNE		
AÑO	PESO FRESCO Kg ha⁻¹ año⁻¹	PESO SECO Kg ha⁻¹ año⁻¹
1970	117560,6	18796,7
1971	110801,5	17457,8
1972	123607,7	18739,1
1973	115304,7	18931,6
1974	117911,7	19046,4
1975	130639,4	20996,8

PRODUCCION ANUAL MEDIA de dos cuadrados latinos de 9 × 9 parcelas de un metro cuadrado de superficie cada una. Se observa una producción sumamente elevada, por encima de las buenas cosechas de pratenses (unas 15 Tm ha^{-1} $año^{-1}$) y ello, sin tener en cuenta la producción neta del rastrojo, ni la de las partes subterráneas, que elevarían la producción total neta por encima de las 30 Tm ha^{-1} $año^{-1}$, en el límite de las producciones potenciales máximas calculadas. La producción más elevada, para uno de los dos campos promediados, corresponde al año 1975 con 21465,7 kg ha^{-1} $año^{-1}$; la mínima, al año 1971, con 17234,8 kg ha^{-1} $año^{-1}$ de peso seco obtenido por siega.

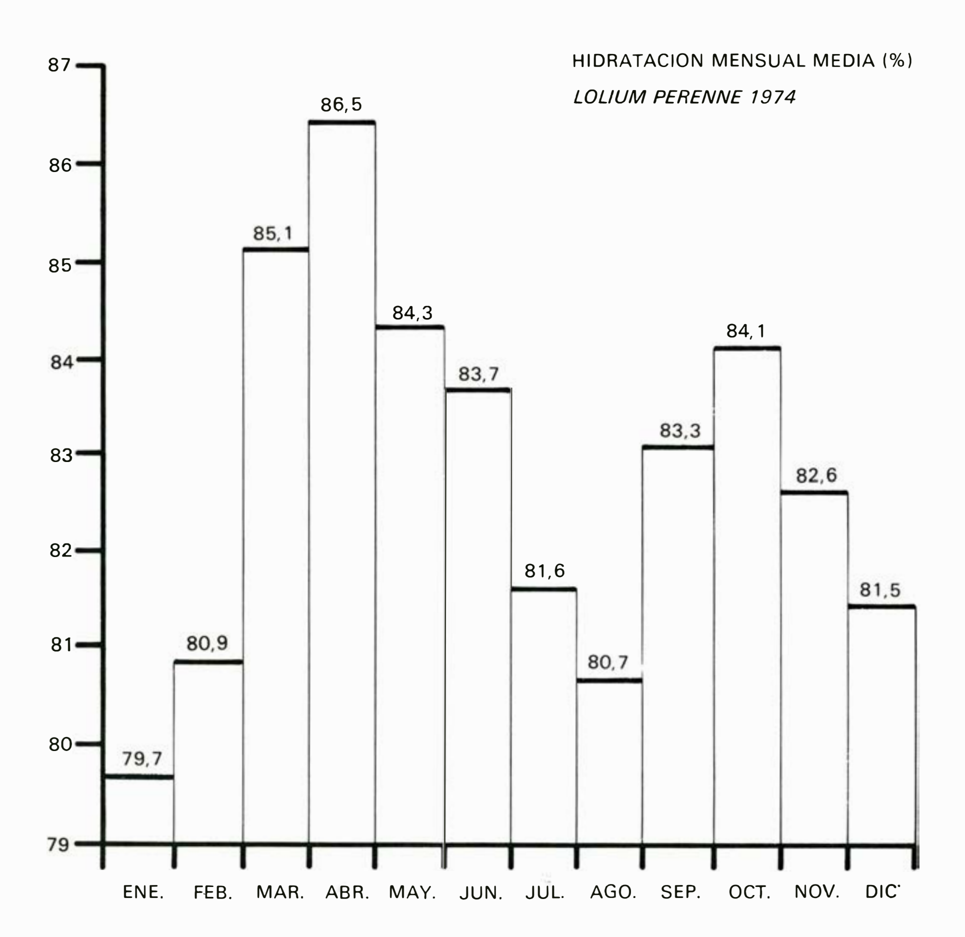
desenvuelven en condiciones deficitarias de nitrógeno y de agua y, por ello, sólo rinden una pequeña fracción de su potencialidad para los mismos esfuerzos humanos y condiciones de clima.

Dentro del marco de los condicionamientos económicos, en la actualidad, los sistemas agrícolas tienden a eliminar la limitación originada por el déficit nitrogenado. La utilización de estos programas obliga también a suministros apropiados de agua y con frecuencia se originan, por ello, otras consecuencias que afectan al funcionamiento del vegetal. Así, por ejemplo, se modifica el cociente partes subterráneas/partes aéreas, se favorece el desarrollo vegetativo en perjuicio del reproductor (floración y fructificación), se altera el reparto de asimilados entre los órganos exportadores e importadores, etc.

Los actuales programas de selección de nuevas variedades tienen muy presentes estas circunstancias así como las alteraciones que se originan en los medios enriquecidos. Estos parámetros también se ponderan cuidadosamente en la elaboración de modelos simulados que integran la producción y las fases de crecimiento y del desarrollo.

El Departamento de Fisiología Vegetal de la Facultad de Biología de la Universidad de Barcelona mantiene sin interrupción, desde 1960 hasta la fecha, estudios experimentales de campo, los cuales afectan, sobre todo, al cultivo de especies de gramíneas pratenses *(Lolium perenne, Dactylis glomerata* y *Festuca elatior (arundinacea),* pertenecientes al grupo fotosintético C_3. Dichos trabajos han constituido, en parte, una colaboración internacional que fue dirigida por el Dr. Gerhard Rappe (Vassmolösa, Suecia) hasta su fallecimiento en 1974.

Los fines perseguidos con estos estudios han sido diversos, pero, al principio, se concentraron en conocer las características y averiguar las causas de la depresión estival de la producción que estas especies presentan en los climas templados y mediterráneos. Simultáneamente, se trató de resolver las posibles fluctuaciones rítmicas subestacionales, estacionales, anuales y supranuales del crecimiento y de la producción de materia fresca y seca (Caballero, Ubach, Berbel, Morales, 1968 y Caballero, 1969).



EL CONTENIDO DE AGUA de la materia viva es un indicador indirecto de la actividad vital. En la figura se exponen las fluctuaciones estacionales del grado de hidratación medio mensual de *Lolium perenne.* Este histograma se presenta claramente bimodal, con un máximo principal en primavera y otro en otoño, separados entre sí por la depresión estival. Resulta, así, evidente, que las épocas de mayor productividad se corresponden, de modo muy preciso, con la mayor hidratación de la materia vegetal, ya que se encuentra alta correlación entre la hidratación mensual media y la producción mensual. Obsérvese el parecido entre este histograma y la curva de producción.

Durante los últimos años, sobre todo desde 1970, el mayor interés se ha volcado en conocer y tratar de dominar los parámetros que permitan llegar a maximalizar la productividad de estas especies en clima mediterráneo. Esta última fase en la que aquí nos concentramos, comprende los siguientes aspectos sobresalientes, que influyen en la producción estacional y anual:

a) metodología de cosechado utilizada para medir la producción; b) régimen de irrigación; c) régimen de abonado; d) capacidad fotosintética del prado, y e) eficiencia de la conversión energética.

Las especies ensayadas, anteriormente mencionadas, fueron cultivadas en tres localidades diferentes: "Campos de Deportes" (1960-1966), "Palacio de Pedralbes" (1964-1967) y "Campos Experimentales" (1967 - hasta la actualidad), todas ellas situadas en la zona de Pedralbes, al Oeste, y en las proximidades, de la ciudad de Barcelona.

Las pratenses ofrecen cualidades muy favorables para este tipo de estudios por la facilidad que presentan para medir y seguir el curso de su producción por el método de siegas sucesivas, basado en la capacidad del rastrojo remanente para continuar creciendo por rebrote o ahijado.

Para las tres especies se utilizaron grupos de parcelas mantenidas en régimen

homogéneo de cultivo (riego, abonado, desherbado, etc.). El césped desarrollado en cada parcela individual se cortó con tijera, siempre a la misma altura de 3 cm sobre el suelo, y siguiendo una frecuencia de veintiún días. Dentro de cada grupo de parcelas pertenecientes a la misma especie se mantuvo un cosechado en sucesión cíclica constante.

Uno de los parámetros que intervienen significativamente en la producción de los prados de cosecha continua es la propia cadencia de los cortes. Dadas las características sigmoidales de las curvas generales de crecimiento, es necesario hallar, para cada variedad ensayada y zona climática determinada, una intermitencia tal que sea suficientemente amplia como para alcanzar la fase de crecimiento rectilíneo máximo y hacer mínimo el gasto de energía mecánica de recolección, por equivalente de energía alimenticia obtenida, y suficientemente frecuente como para que el vegetal no llegue a entallar, con lo que se provocaría una pérdida apreciable en la producción total.

La frecuencia de tres semanas, así como la altura del corte, fueron seleccionadas por ensayos previos y resultaron muy adecuados para las condiciones de Barcelona; sobre todo, cuando se impusieron los programas de fuerte abonado nitrogenado y de riegos frecuentes y controlados.

La defoliación sufrida por la siega de césped de zona templada, en el momento del alargamiento del tallo, durante la floración o espigado, suele provocar una inhibición duradera del desarrollo de los nuevos rebrotes, con lo cual toda la producción subsiguiente puede quedar modificada y, con frecuencia, muy perjudicada. Con el período de tres semanas la planta no puede llegar a encañar, ni siquiera en la época más propicia para ello (abril-mayo) y el perjuicio queda así soslayado.

La adecuación de esta metodología con dicho ritmo de corte, queda confirmada a la vista de las elevadas producciones obtenidas y por no presentarse síntomas de degradación de los prados en el transcurso de los años de experimentación.

Las determinaciones realizadas con el material cosechado fueron de peso fresco, es decir, sin haber eliminado el agua del vegetal, por unidad de superficie cultivada, obtenido directamente en el campo después de cada corte, y de peso seco tras desecación en estufa a 100-105º C.

Con los pesos obtenidos de las parcelas en sucesión y referidos a metros cuadrados de superficie de cultivo se construyeron gráficos como los que se incluyen, cuyos perfiles anuales expresan las variaciones de la tasa de producción en el tiempo (gramos de peso seco por metro cuadrado y por períodos de veintiún días) para cada fecha de corte.

Con objeto de conocer la biomasa no incluida en los cortes (peso de la raíz y del rastrojo), se ha obtenido periódicamente muestras de suelo, incluyendo raíz y rastrojo, con ayuda de un taladro cilíndrico que se clava hasta una profundidad máxima de 20 cm.

A partir de 1964, se empezaron a utilizar agrupaciones de parcelas dispuestas en cuadrado latino con el fin de lograr mejor significación estadística en los resultados. En los campos mayores, de ochenta y una parcelas, se obtuvieron 27 datos primarios semanales por campo.

Las parcelas estuvieron rodeadas de marcos cultivados con la misma especie para eliminar el llamado efecto de borde y en todos los campos se mantuvieron, además, distintas parcelas con fines específicos: conocer la época de floración, efectos del abonado, determinación de la actividad fotosintética y respiratoria, medida de parámetros físicos (suelo - aire), etc.

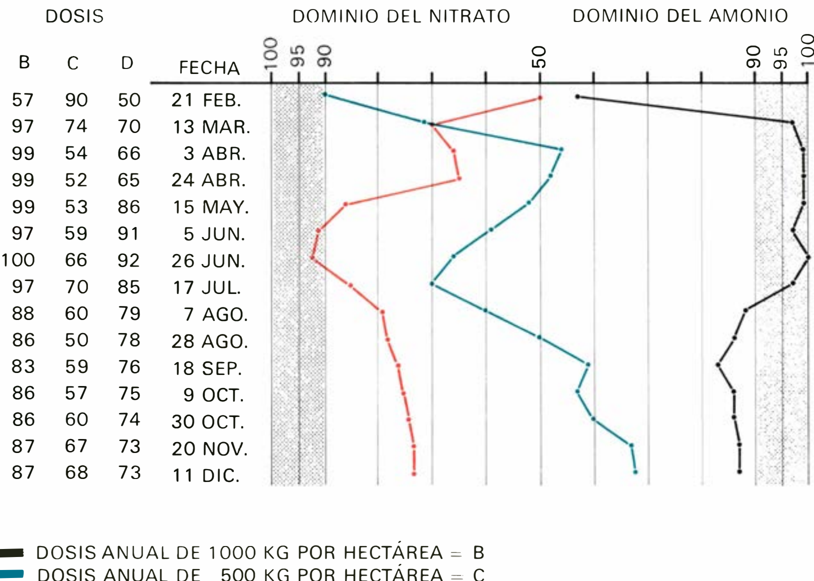
En cuanto al riego, se procuró mantener el suelo constantemente húmedo; los dos primeros campos se regaron usando mangueras o por inundación. Después de comprobarse la importancia de un adecuado régimen de irrigación, desde 1969 se habilitó una red de aspersión para la distribución homogénea y frecuente del agua.

Los programas de abonado variaron se-

NIVEL DE SIGNIFICACIÓN DE DIFERENCIAS DE PRODUCCIÓN - NITRATO Y AMONIO - A IGUAL DOSIS

CAMPO DE 14 PARCELAS

PESO SECO



B	C	D	FECHA
57	90	50	21 FEB.
97	74	70	13 MAR.
99	54	66	3 ABR.
99	52	65	24 ABR.
99	53	86	15 MAY.
97	59	91	5 JUN.
100	66	92	26 JUN.
97	70	85	17 JUL.
88	60	79	7 AGO.
86	50	78	28 AGO.
83	59	76	18 SEP.
86	57	75	9 OCT.
86	60	74	30 OCT.
87	67	73	20 NOV.
87	68	73	11 DIC.

REPERCUSION ESTACIONAL de la nutrición nitrogenada. En abscisas, tiempo; en ordenadas, nivel de significación (%) de las diferencias de producción obtenidas a igual dosis de abonado con fertilizante amoniacal y nítrico. Los colores negro, azul y rojo se corresponden con las dosis de 1000, 500 y 250 kg N ha^{-1} $año^{-1}$. Resulta claramente explícito que las altas concentraciones de nitrógeno (1000 kg N ha^{-1} $año^{-1}$) producen diferencias enormemente significativas de producción a favor del amonio; las bajas (250 kg N ha^{-1} $año^{-1}$), a favor del nitrato, y, las intermedias (500 kg N ha^{-1} $año^{-1}$) originan producciones cuyas diferencias oscilan siempre en el campo de significación inferior al nivel del 90% (diferencias no significativas). Sin embargo, en las tres curvas, se aprecia un ascenso hacia el dominio del amonio en los meses de insolación más baja y un descenso al dominio del nitrato en los meses de mayor iluminación, aunque presentándose un efecto de retraso según las dosis. Así, la mayor tendencia al beneficio del abonado amoniacal se observa en abril. para las dosis bajas, mayo para las intermedias y junio, para las altas; la acción benéfica del nitrato se sitúa en junio, julio y agosto, respectivamente. Ello está en consonancia con el supuesto de que en condiciones de gran luminosidad el ion nitrato puede reducir la fotorrespiración e incrementar, por consiguiente, la fotosíntesis neta actuando como sumidero de energía fotosintética.

gún los años y los campos de parcelas. A lo largo de las series de ensayo, se pasó de los niveles bajos de nitrógeno habituales en praticultura (100 a 250 kg N por hectárea y año) a cifras masivas muy superiores a las máximas recomendadas (1000 a 2000 kg de nitrógeno por hectárea y año).

En el "Campo de Deportes", en donde se habían plantado *Dactylis* y *Festuca*, en 1961 no se suministró fertilizante alguno. A partir del año siguiente, y a la vista de los bajos rendimientos obtenidos (entre 4 y 5 Tm peso seco por hectarea y año) se utilizó siempre abono mineral ($N:P_2O_5$: K_2O=1:1:1) repartido en fracciones iguales, siguiendo el ritmo de los cortes, después de cada ciclo. En 1962, 100 kg N; en 1963, 190 kg N y, a partir de 1964, 250 kg N. Durante ese mismo año se alcanzaron rendimientos especialmente llamativos (de 16 a 20.5 Tm) probablemente por haberse realizado simultáneamente un cuidadoso régimen de riego, el cual, al desmerecer en los dos años siguientes, 1965 y 1966, disminuyó la productividad, manifestando los prados síntomas de degradación evidente (de 10 Tm a 4,5 Tm).

En el campo "Palacio de Pedralbes" (1964-1967), se continuó aplicando igual dosis de 250 kg N en las tres especies y se volvió a comprobar la importancia primordial de una frecuente aplicación de riego, sobre todo durante la época calurosa.

En los "Campos Experimentales" se controlaron el riego y el abonado nitrogenado con ayuda de cajas lisimétricas y análisis de planta y suelo. Los resultados obtenidos (Simón, 1972) y las recomendaciones y conclusiones al respecto de Klapp (1959) y de Aspinal (1961) referentes a la influencia del abonado y de su forma de aplicación en las depresiones de la producción estival, nos decidieron a elevar las dosis (640 kgN) llegándose después, en los últimos años, a cantidades muy considerables (1000 a 2000 kg N).

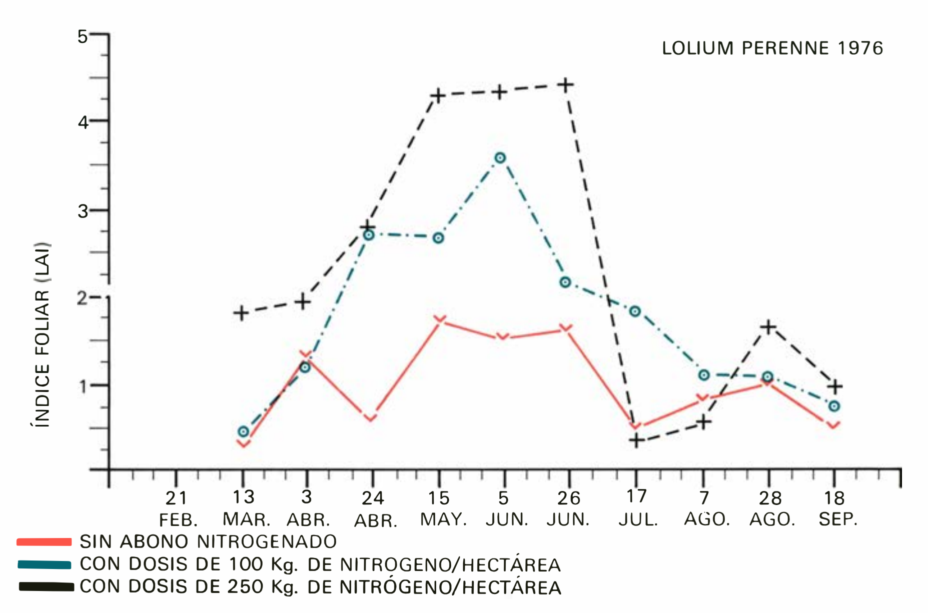
Estas cifras finales, en apariencia tan exageradas, se impusieron al comprobarse que el régimen intenso de aspersión que se mantenía provocaba algunas pérdidas por lavado y que, en las épocas de mayor crecimiento del césped, la disolución del suelo quedaba exhausta de nitrógeno (Simón 1972).

La combinación racional de abonados nitrogenados intensos, fraccionados adecuadamente durante todo el año, y los riegos por aspersión, muy frecuentes durante el verano, llegaron a proporcionar normalmente producciones de corte próximas a las 20 Tm peso seco por hectárea y año (en ocasiones rebasadas).

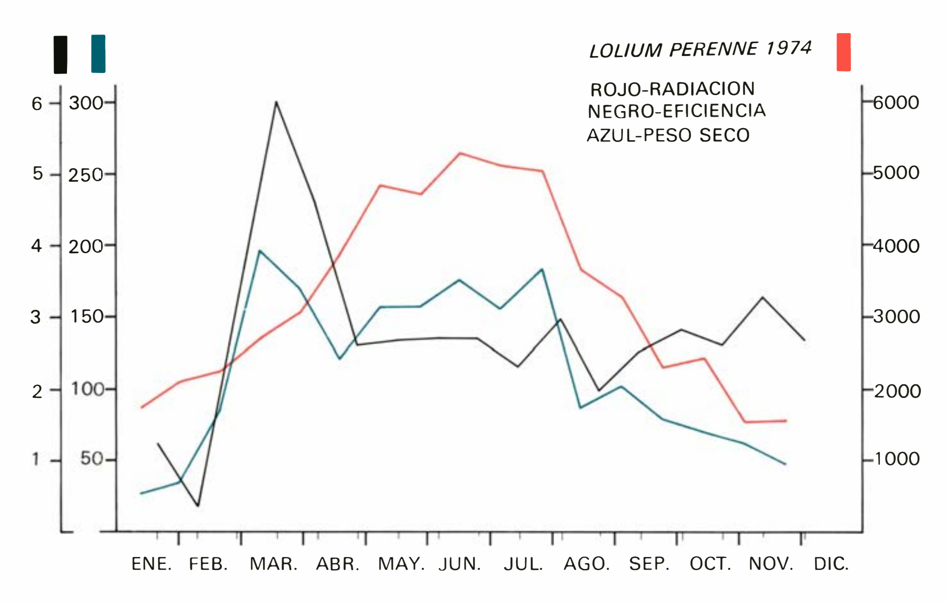
Las clásicas depresiones durante la época calurosa quedaron muy disminuidas, pasándose así de unas distribuciones anuales de producción de perfil bimodal o trimodal a otras multimodales, de tasa de producción bastante mantenida.

La comparación de los gráficos de la página 90 de los primeros "Campos de Deportes" y los pertenecientes a las finales de "Campos Experimentales", en la página 91 permite apreciar claramente las diferencias indicadas en los perfiles. También las Tablas de la página 92 destacan las diferencias de producción de corte en ambas series de ensayos.

En cuanto a las producciones anuales de forraje seco en prados cultivados de



EL INDICE FOLIAR (LAI) indica la relación entre la cobertura foliar de una vegetación (superficie total de una de las caras de las hojas) y la superficie de suelo que ocupa. En la gráfica se refleja la variación del LAI (expresada en metros cuadrados de hoja por metro cuadrado de suelo) a lo largo del año, para plantas sin abonado nitrogenado *(rojo)* y con dosis de 100 *(azul)* y 250 *(negro)* kg N ha^{-1} $año^{-1}$. Se aprecia cómo aumenta el índice foliar en el abonado y cómo es más alto en las épocas de mayor producción, en las cuales queda comprendido entre 3 y 5 que son, precisamente, los valores que se consideran óptimos para un mayor aprovechamiento de la luz solar.



OSCILACIONES TEMPORALES de la radiación fotosintéticamente activa *(rojo)* expresada en calorías por centímetro cuadrado (1 caloría aproximadamente 4,18 julios), el peso seco *(azul)* en gramos por metro cuadrado (1 gramo de materia seca rinde unos 0,0176 megajulios) y la eficiencia fotosintética *(negro)* en %, para un solo grupo de parcelas. Se observa claramente como las mayores eficiencias se obtienen en primavera, con un gran incremento del peso seco y radiaciones relativamente bajas, mientras que en la época de depresión estival, con insolación alta, la pérdida de producción, debida seguramente a efectos de la temperatura excesivamente elevada, repercute en eficiencias claramente bajas. (Datos obtenidos de un estudio de J. Mª Recasens, del Departamento de Fisiología Vegetal de la Facultad de Biología de la Universidad de Barcelona).

zona templada, se considera ya las cifras de 15 Tm por hectárea y año como rendimientos muy elevados. Varios autores (Tainton 1968, Cooper 1970, Stern 1975, Moos 1976) han calculado las producciones máximas potenciales; las cifras que se obtienen son del orden de las 30 a 35 Tm por hectárea y año, es decir, el doble de las buenas cosechas reales.

Si bien hemos sobrepasado, en ocasiones el límite de las 20 Tm por hectárea y año, hay que tener en cuenta que esta producción se refiere sólo a la porción vegetal obtenida por corte. Si se incluye la producción neta que supone el rastrojo remanente de las siegas, se debe sumar como mínimo de 3 a 4 Tm por hectárea y año. También es necesario tener presente que, durante el período vegetativo anual, la producción de raíz supone otra cifra importante que eleva, en realidad, la neta total en 4 o 5 Tm por hectárea y año. Estas cifras son bajas debido al riego, al abonado y a los cortes aplicados con elevada frecuencia. De todo ello resulta una producción real máxima del orden de 30-32 Tm de peso seco por hectárea y año, en el límite de la producción potencial máxima.

Los perfiles anuales del recrecimiento de las tres especies de *Dactylis, Festuca* y *Lolium* ofrecen mayores semejanzas entre sí, dentro de un mismo año e incluso dentro de campos distintos, que los de una misma especie en años sucesivos. Ello destaca sobre todo en los ensayos que utilizaron niveles bajos de nitrógeno y riegos poco frecuentes (acción preponderante de los parámetros ambientales o climáticos).

Aparte de los máximos principales a que nos hemos referido anteriormente, los perfiles de producción de las variedades ensayadas presentan una serie de depresiones y máximos relativos que se distribuyen de forma aparentemente periódica. Mediante técnicas de reconocimiento y valoración de periodicidades en series temporales, se puede inferir la existencia de un ritmo de unos 35 días de período que parece interaccionar con ritmos de naturaleza lunar, de modo que aparecen fluctuaciones de menor orden que coinciden aproximadamente con la semana y el semimés lunares. Estos máximos relativos de crecimiento se revelan mucho más patentes en el peso fresco de la producción que en el peso seco; en gran parte, pues, lo que aumenta realmente es el contenido medio de agua en el vegetal, lo que da la idea de una especie de "marea" en el interior de la planta.

Como antes se ha establecido, los factores limitantes que más comúnmente inciden en la producción de pratenses son, indiscutiblemente, la disponibilidad hídrica y el acervo de nitrógeno asimilable.

No existe ninguna duda sobre la existencia de una relación directa entre el contenido de agua de un vegetal y su propia actividad. Ello es debido a las múltiples e imprescindibles funciones que esta sustancia verifica en la planta, suministrando el grado de hidratación necesario para la estructuración y fisiología de las macromoléculas hidrófilas del protoplasma (como proteínas, ácidos nucleicos y glúcidos), aportando seno ambiente en el que transcurre la mayor parte de las reacciones químicas vitales (algunas de ellas solamente posibles en disolución), actuando como solvente de los electrólitos y de las moléculas polares, catalizando diversas reacciones e interviniendo directamente en otras, como producto intermediario (hidrólisis y condensaciones) o bien como producto inicial (fotosíntesis) o final (respiración), aparte de las misiones puramente físicas como las de mantener la turgencia celular, colaborar en el transporte de nutrientes y asimilados.

Por ello no es de extrañar que para obtener buenos rendimientos se haga imprescindible mantener el césped constantemente húmedo, y que esta necesidad se manifieste mucho más imperiosa cuando se trata de lograr una distribución anual satisfactoria de la producción. En nuestros Campos Experimentales se ha comprobado empíricamente que es fundamentalmente precisa una irrigación con dos días de intermitencia máxima para pretender una cosecha estival suficiente.

Para comprobar el efecto de la hidratación en la producción, se ha diseñado un ensayo en el que se ha procedido al corte de parcelas cada sesenta minutos, entre las ocho y las dieciocho horas solares durante dos años sucesivos. Con cada muestra obtenida de este modo se ha determinado el peso fresco y el peso seco. Estos dos datos permiten conocer, para un tiempo determinado, el porcentaje de peso seco en la muestra o bien el parámetro inverso, la hidratación del material. Los resultados obtenidos conducen a conclusiones interesantes. En efecto, el grado de imbibición acuosa de la planta presenta una oscilación anual máxima (datos referentes a 1974) de alrededor del 10 por ciento, con má-

LOLIUM PERENNE

	MAXIMOS DE PRODUCCION					MINIMOS DE PRODUCCION				
AÑO	MES	PESO SECO $g\ m^{-2}\ d^{-1}$	EQUIVALENTE ENERGETICO $MJ\ m^{-2}\ d^{-1}$	RADIACION FOT. ACTIVA $MJ\ m^{-2}\ d^{-1}$	EFICIENCIA %	MES	PESO SECO $g\ m^{-2}\ d^{-1}$	EQUIVALENTE ENERGETICO $MJ\ m^{-2}\ d^{-1}$	RADIACION FOT. ACTIVA $MJ\ m^{-2}\ d^{-1}$	EFICIENCIA %
1972	ABRIL	13,81	0,24	4,80	5,00	SEPTIEMBRE	2,61	0,05	6,79	0,74
1973	ABRIL	16,36	0,29	5,44	5,24	SEPTIEMBRE	2,83	0,05	9,43	0,53
1974	ABRIL	11,18	0,20	5,23	3,73	SEPTIEMBRE	3,23	0,06	6,66	0,85
1975	MAYO	19,09	0,34	7,76	4,32	AGOSTO	1,67	0,03	8,72	0,34
1976	ABRIL	12,21	0,21	7,57	2,84	AGOSTO	2,62	0,05	9,50	0,49

EFICIENCIA FOTOSINTETICA de *Lolium perenne* en los máximos y mínimos de producción (1972-1976). Este parámetro relaciona la energía acumulada en la materia vegetal respecto a la energía solar absorbida por la planta (radiación fotosintéticamente activa, aproximadamente el 50 % de la que llega al vegetal). La proporción entre la energía luminosa aprovechable en la fotosíntesis (PAR) y la que es almacenada en la materia orgánica suele quedar próxima al 2 %, como valor promedio anual. Los máximos de eficiencia quedan comprendidos entre 2.8 y 5.2 % y coinciden con los máximos de producción. Los mínimos, a su vez, oscilan entre 0,34 y 0,85, que corresponden a las depresiones estivales.

ximos en abril (cerca del 87,5 por ciento) y mínimos en enero (77,5 por ciento). Esta situación permite agrupar los meses de este año en dos categorías según la hidratación vegetal: meses fisiológicamente secos (enero, agosto, febrero, diciembre y julio, ordenados de menor a mayor contenido de agua) y meses fisiológicamente húmedos (abril, marzo, mayo, octubre, junio y septiembre, con hidratación decreciente), con un mes de transición (noviembre).

Resulta evidente que las épocas de mayor productividad se corresponden, de modo muy preciso, con las de mayor proporción hídrica en la materia vegetal y que los mínimos de crecimiento están en consonancia con las épocas de mayor sequedad fisiológica. De todo ello se deduce que la curva que relaciona la hidratación mensual media con el tiempo se presenta claramente bimodal, con un máximo principal en primavera y otro en otoño, lo cual se correlaciona muy exactamente con el perfil de producción anual. Se puede concluir, por consiguiente, que producción e hidratación sufren una evolución paralela.

Respecto a la nutrición nitrogenada aunque, en principio, en suelos neutros (con pH próximo a 7), equilibrados, amortiguados y con carbonato cálcico no parece probable, a priori, que se presenten diferencias entre la absorción y la asimilación del nitrato y del amonio, existen datos que prueban (Hofstra, Koch-Bosma, 1970), en especies C_3 como el tomate, que con radiación solar baja se obtiene una mayor producción de materia seca suministrando nitrógeno amoniacal; por contra, con insolación elevada, el nitrato resulta más eficaz. Del mismo modo, en cultivos hidropónicos de claveles (Green y Holley, 1974), se ha encontrado que la producción se incrementa a niveles comparables a los de plantas C_4 si se utiliza nitrato como única fuente de nitrógeno en los períodos de radiación solar más elevada. La reducción del nitrato a materia orgánica aumentaría la asimilación neta de CO_2, disminuyendo la fotorrespiración y utilizándose, por consiguiente, el nitrato como sumidero de energía fotosintética.

Para dilucidar la repercusión estacional de la nutrición nitrogenada en los aspectos diferenciales atribuibles al suministro en forma aniónica (nitrato) o catiónica (amonio), hemos acondicionado (Gil, Caballero, Guadilla, 1976) dos campos, uno de 14 parcelas de 2 m² de superficie cada una (para ensayos de dos repeticiones de seis dosis de abonado y un control) y otro de 81 parcelas de 1 m² (para nueve repeticiones de ocho dosis y un testigo). Se han considerado como más oportunos los niveles de 250, 500 y 1000 kg por hectárea y año, tanto para el NH_4^+ como para el NO_3^-.

Del estudio computacional de los dieciocho parámetros establecidos con los datos obtenidos se desprenden las siguientes conclusiones:

a) Los distintos niveles de abonado ensayados son insuficientes, por sí solos, para eliminar totalmente el descenso estival de producción. Por otro lado , parece improbable que este efecto se pueda lograr mediante fertilizaciones nitrogenadas más intensas, ya que las dosis más elevadas que hemos suministrado llegan a alcanzar los niveles de saturación, a pesar del enérgico régimen de riego a que han sido sometidas las parcelas. Sin embargo, se ha apreciado que el abonado conduce a un adelanto en el tiempo en la obtención de los máximos de cosecha; por ello, en la época de depresión estival, la producción de las parcelas control, no fertilizadas, puede sobrepasar a la que se obtiene con las distintas dosis de fertilización, produciéndose una inversión de la situación normal en que las parcelas abonadas son más productivas que las testigo. Además, el abonado conduce a un incremento en el peso seco y en el peso fresco, pero en este último caso el rendimiento es mayor, lo que equivale a decir que la fertilización incrementa la hidratación; este efecto ha sido más intenso para la forma nítrica que para la forma amoniacal.

b) Sin originar prácticamente repercusiones en la linealidad de respuesta, las altas concentraciones (1000 kg N por hectárea y año) de abonado nitrogenado determinan diferencias significativas de producción a favor del amonio, por lo menos en la mayor parte del ciclo anual; las bajas concentraciones (250 kg N por hectárea y año) originan diferencias significativas a favor del nitrato, y, finalmente, las dosis intermedias (500 kg N por hectárea y año) dan lugar a producciones muy similares, que no difieren significativamente.

Por lo tanto, el distinto efecto de la dieta nítrica y amoniacal depende más de la concentración que de la radiación solar. Existe, sin embargo, un desplazamiento relativo a favor del amonio con insolaciones bajas y del nitrato con fuerte insolación (¿en relación con la fotorrespiración?), que estaría de acuerdo con los datos bibliográficos acerca del mismo efecto en cultivos hidropónicos. Estos desplazamientos son tanto más retardados cuanto mayor sea la concentración del fertilizante. Si a todo ello sumamos el hecho de que, en la depresión estival, se produce una inversión de la relación dosis/efecto, queda bastante claro que existen diversos puntos de actuación para la regulación estacional de las producciones, mediante el ensayo de diversas proporciones y dosis de los iones amonio y nitrato según la etapa de crecimiento y época del año.

Como ya se ha comentado anteriormente, la asimilación fotosintética del anhídrido carbónico atmosférico por los tejidos verdes de la planta constituye el proceso principal en la producción de materia orgánica. Por el contrario, los procesos respiratorios suponen consumo de asimilados con desprendimiento de dicho gas, por lo que el balance de ambas actividades, la fotosíntesis neta, decide primordialmente la producción.

Por ello se consideró interesante estudiar en condiciones de campo el intercambio gaseoso asociado a la fotosíntesis y a la respiración, utilizándose la metodología habitual (Leafe, 1972) para estos casos. Sobre todo, pareció necesario realizarlo durante períodos decisivos tales como las depresiones estivales.

Una forma corriente para expresar la capacidad fotosintética neta consiste en referir la cantidad de CO_2 fijado por unidad de superficie foliar (la de la cara que recibe la luz) y por unidad de tiempo (mg CO_2 por decímetro cuadrado y hora) y las cifras que se encontraron en parcelas de *Lolium perenne* fueron las normales, entre 18 y 32 mg, según los casos, y siempre que no fueran limitantes la intensidad de luz solar incidente y el suministro de agua (Caballero, Guadilla Gil, Simmoneau, 1976).

En ocasiones, como por ejemplo después de una siega de elevada producción de peso seco, se registraron balances negativos en el rastrojo, al dominar la respiración sobre la fotosíntesis, que rápidamente se compensaron y volvieron a ser positivos en los días inmediatos, al reanudarse la rebrotación y el crecimiento.

Sin embargo, y en contra de lo que se podía esperar, incluso durante las depresiones de la producción en agosto, se encontraron cifras positivas notables. Parece ser que la planta experimenta durante el estío adaptaciones a radiación solar y temperatura más elevadas que mejoran o evitan reducciones importantes de su eficacia fotosintética. Se deduce así que no es precisamente el descenso de la capacidad fotosintética lo que provoca el bloqueo de la producción.

Es un hecho conocido que la raíz del

césped es muy sensible y queda fácilmente perjudicada por las temperaturas elevadas del suelo. Por encima de los 38° C el aparato radicular puede sufrir daños irreparables. La circunstancia de que las capas superficiales del suelo rebasen estas temperaturas durante los días más calurosos y calmados, podría ser causa y explicación del perjuicio que experimenta el césped en tales ocasiones. Precisamente, los frecuentes riegos por aspersión no permitirían los excesos térmicos, evitando con ello o, por lo menos, reduciendo las depresiones acentuadas.

De todos modos, no debe olvidarse que los elevados niveles de abonado y los riegos abundantes favorecen un desarrollo superficial y más limitado del conjunto radicular, lo cual aumentaría el peligro de lesiones térmicas, por ser precisamente las capas más superficiales del suelo las que alcanzan temperaturas más altas en las horas en que la insolación sea mayor.

Cuando la actividad fotosintética se expresa en relación con toda la superficie foliar (una sola cara) de la vegetación que ocupa una determinada superficie del suelo, se encuentran correlaciones muy interesantes con la producción estacional. Los índices foliares (LAI) que representan la relación entre la superficie foliar total (una cara) de la vegetación y la del suelo ocupado por ésta pueden suponer cifras de 10 a 12 en prados abonados no segados de *Lolium perenne* durante las épocas en las que se alcanza el mayor crecimiento. Los máximos encontrados por nosotros son de 3 a 5 y coinciden con las siegas de máximas producciones. Es sabido que, precisamente estos índices, son los que proporcionan en general mayores rendimientos en la vegetación y en nuestros ensayos vendrían condicionados por los periodos de tres semanas de siega que evitarían las cargas foliares excesivas que mermarían la eficacia *(véase la ilustración de la página 94).*

Se ha podido comprobar que, desde finales de la primavera, el césped de *Lolium* presenta, en relación con su actividad fotosintética, unos niveles de saturación luminosa muy altos (entre 0,8 y 1,0 calorías por centímetro cuadrado y por minuto).

Se admite que, de toda la energía solar incidente que mide un radiómetro, sólo son activas en la fotosíntesis las longitudes de onda de 400 a 700 nm, las cuales contienen poco más o menos la mitad de la energía.

Desde el punto de vista de la productividad resulta interesante determinar la llamada eficiencia de conversión energética (luminosa-química), es decir, la relación que existe entre energía luminosa fotosintéticamente activa (PAR) que recibe la planta y la energía que consigue acumular en la materia orgánica producida.

En nuestro caso, para realizar un cálculo aproximado de esta eficiencia se utilizaron las producciones de materia seca correspondientes a las siegas de cada veintiún días, relacionándolas con la radiación solar activa (PAR) recibida durante el correspondiente período de veintiún días de crecimiento. De las integraciones de radiación solar total de onda corta obtenidas con albedómetros Kipp se dedujo la fracción activa (PAR) aplicando los coeficientes estacionales de Britton y Dodd (1976) que suponen entre un 46 y 58 por ciento del total de la energía.

Para la evaluación energética de la materia seca se utilizó la relación 1 g peso seco=4,2 Kcal=17,6 KJ (Lieth 1968).

Las eficiencias medias calculadas en máximos y mínimos de producción de *Lolium,* durante estos últimos años, son elevadas; entre 5,24 y 2,84 por ciento (valor medio 4,2 por ciento) para los máximos y entre 0,85 y 0,34 por ciento (valor medio 0,59 por ciento) en lo que se refiere a los mínimos *(véase la ilustración de la pág. 95).*

En momentos o situaciones determinadas se alcanzan eficiencias que rebasan el 6 por ciento, como en ocasión de los crecimientos de principios de primavera los primeros años de iniciarse el cultivo de un prado o en grupos de parcelas de crecimiento privilegiado *(véase la ilustración de la página 94).*

Estos rendimientos tan altos (de los mayores citados para plantas del C_3) se corresponden con las productividades, también exuberantes, ya comentadas, de los últimos años.

Aunque las posibilidades para mejorar los rendimientos de producción no han quedado, con mucho, agotadas, las producciones netas conseguidas son muy próximas a las potenciales calculadas. Por ello, y mediante la interacción de regímenes variables de riego, abonados nitrogenados con diversas proporciones de amonio y nitrato según la época climática, nivel equilibrado del resto de nutrientes, variación del período de corte atendiendo al LAI y al tiempo, y otras variables fisiológicas, ambientales y de técnica de recolección, es muy probable que las cifras potenciales puedan ser rebasadas.

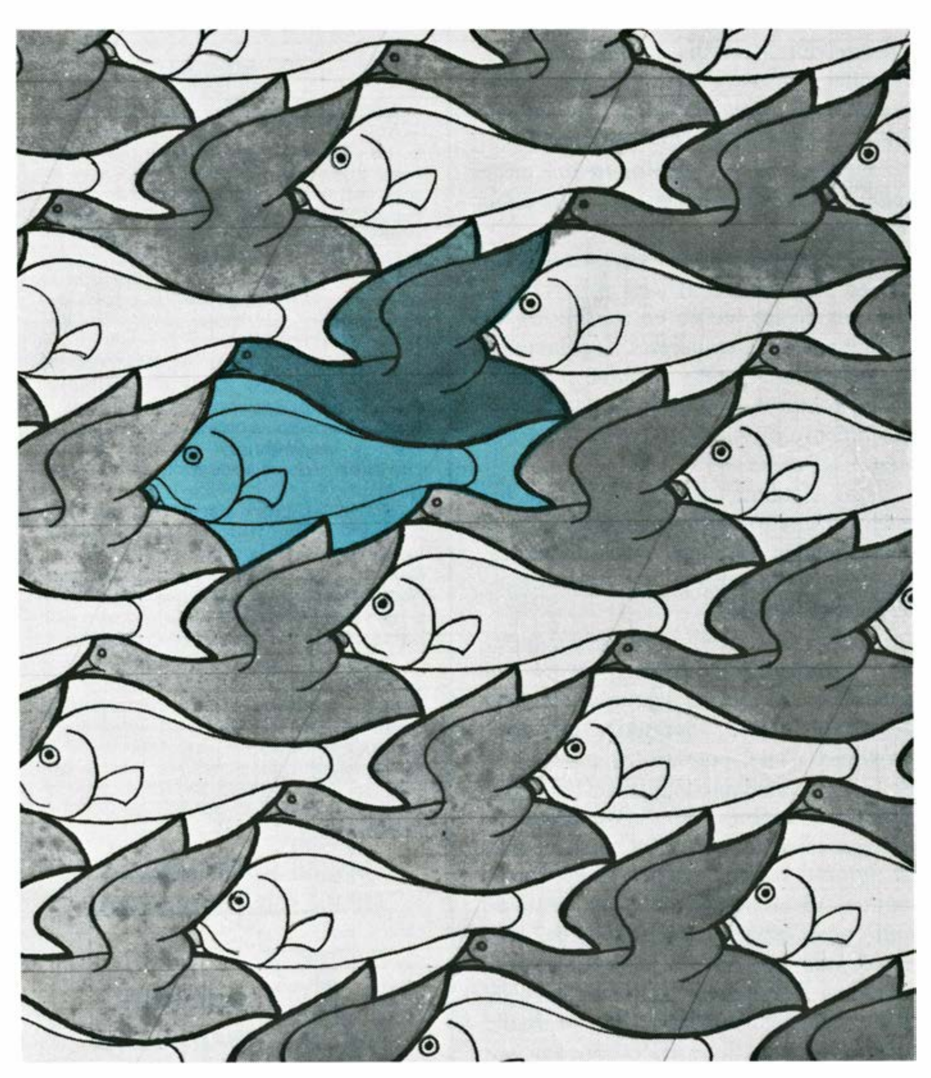
Juegos matemáticos

Extraordinario mosaico no periódico que enriquece la teoría del teselado

Martin Gardner

En agosto de 1975, al final de un artículo de dos partes acerca de cómo embaldosar el plano con polígonos convexos congruentes, prometí un nuevo artículo sobre mosaicos no periódicos. Ahora cumplo dicha promesa y presento por primera vez un importante mosaico no periódico descubierto por el físico-matemático inglés Roger Penrose. Primeramente daré algunas definiciones y presupuestos.

Un mosaico es periódico si en él puede delimitarse una región que se va repitiendo; es decir, una región que "embaldosa" el plano por traslaciones, esto es, cambiando la posición de la región sin girarla ni reflejarla. El artista holandés M. C. Escher debe su fama a sus muchos grabados de mosaicos periódicos con formas que parecen seres vivos. La ilustración inferior es típica; la zona coloreada delimita la región fundamental, que va embaldosando por traslación. Si imaginamos que el plano está cubierto por un papel transparente, en el que esté dibujada cada baldosa, se podrá mover el papel, sin rotaciones, a una nueva posición donde todas las líneas coincidan exactamente, sólo en el caso de que el mosaico sea periódico.



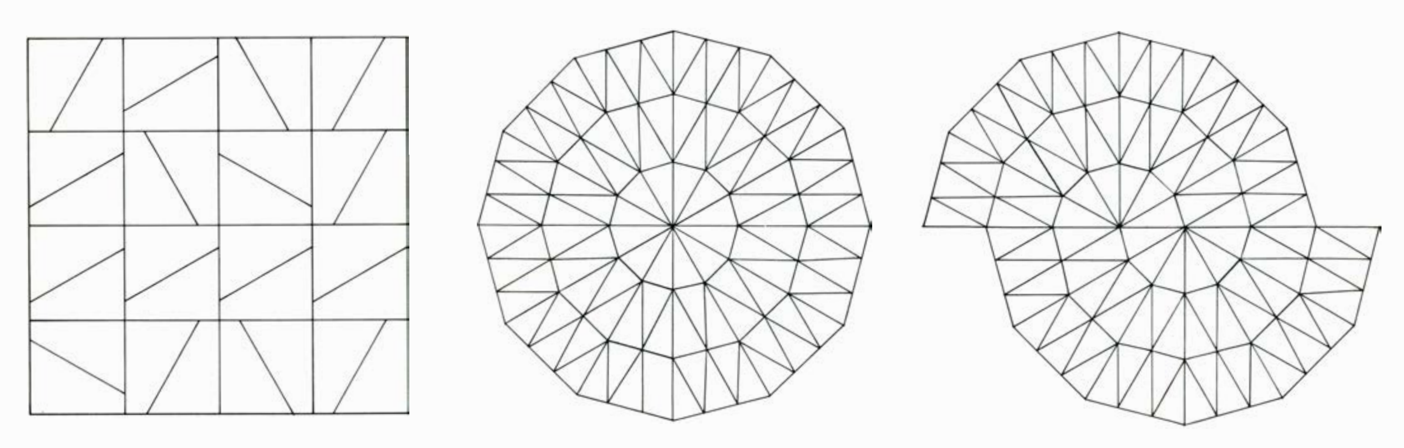
Mosaico periódico construido por M. C. Escher (1938)

Muchas formas, por ejemplo el hexágono regular, embaldosan sólo de manera periódica. Otras embaldosan periódica y no periódicamente. Un tablero de damas se convierte fácilmente en un mosaico no periódico, de triángulos rectángulos isósceles idénticos o de cuadriláteros, simplemente bisectando cada cuadrado, como se muestra arriba, a la izquierda, en la ilustración de la página siguiente, y alterando la orientación para evitar la periodicidad.

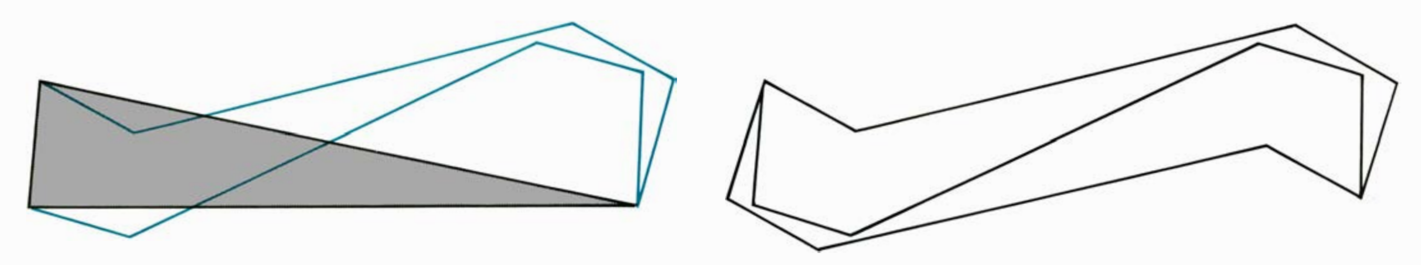
Los triángulos isósceles embaldosan también radialmente, como se muestra en el centro de la ilustración superior y, aunque el embaldosado es muy ordenado, obviamente es no periódico. Tal como Michael Goldberg señaló en su artículo *Central Tessellations* (Mosaicos centrales), publicado en 1955, dicho mosaico puede cortarse por la mitad, y entonces los semiplanos pueden moverse uno o más pasos para dar la forma espiral de un mosaico no periódico como se muestra a la derecha de la misma ilustración. El triángulo se puede deformar de múltiples maneras, substituyendo sus dos lados iguales por líneas congruentes, como se muestra en la ilustración central de la página siguiente. Si los nuevos lados tienen aristas rectas, el resultado es un polígono de 5,7,9,11... aristas, que embaldosan en espiral. La ilustración inferior de la página siguiente muestra un impresionante modelo, obtenido de esta manera a partir de un polígono de nueve lados. Heinz Voderberg fue el primero en llegar a él gracias a un procedimiento complicado; el método seguido por Goldberg para su obtención resulta casi trivial.

En todos los casos conocidos de mosaicos no periódicos de figuras congruentes, la figura embaldosa periódicamente. El dibujo de la derecha, de la ilustración central, muestra como pueden juntarse dos de los eneágonos de Voderberg para dar lugar a un octágono, que embaldosa periódicamente.

Puede obtenerse otro tipo de mosaico no periódico, agrupando baldosas y formando grandes reproducciones de sí mismas. Salomón W. Golomb las llama "reptiles". La ilustración inferior de la página 100, muestra como la forma denominada esfinge embaldosa no periódicamente, dando origen a grandes es-

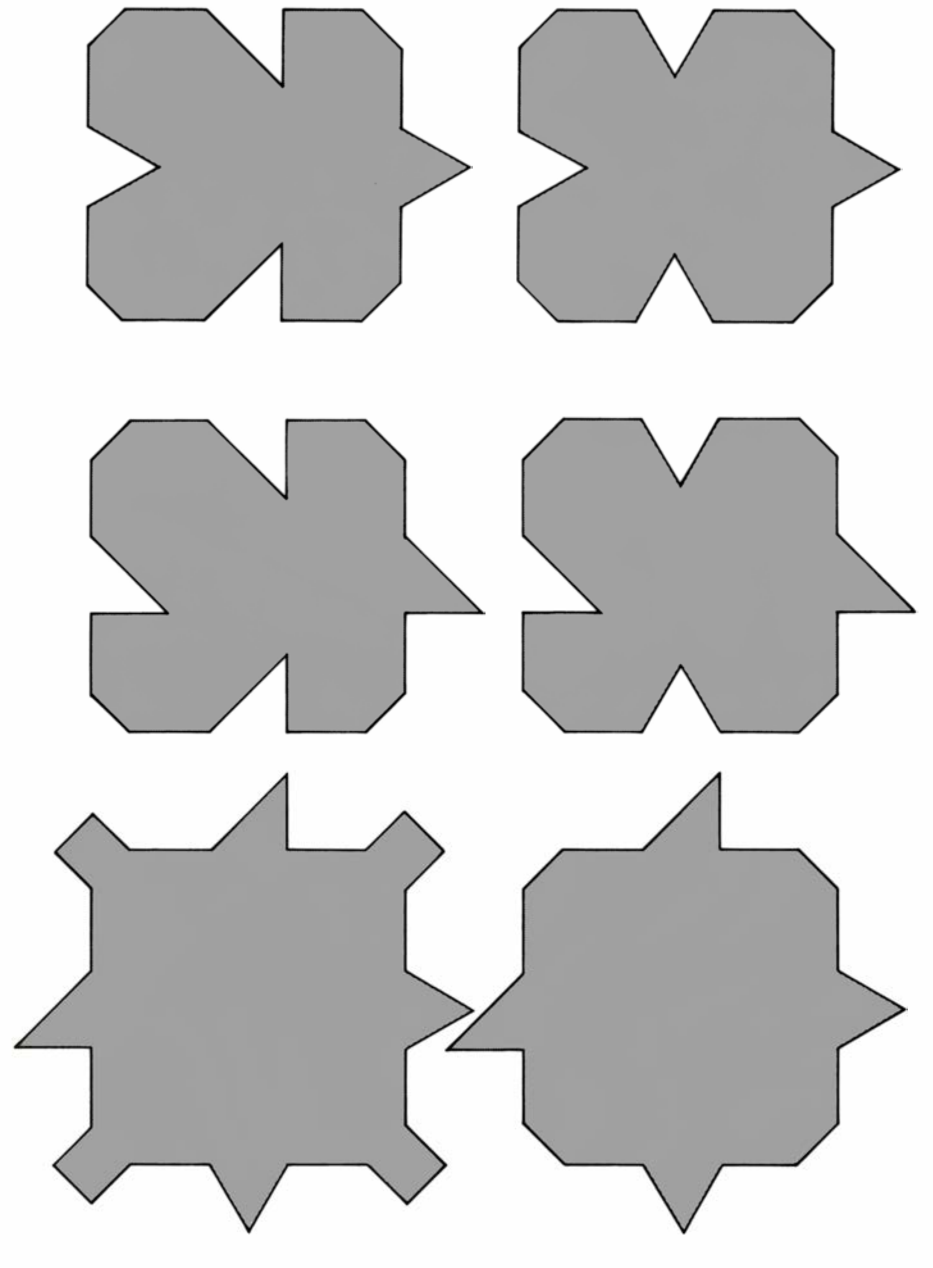


Mosaico no periódico con formas congruentes

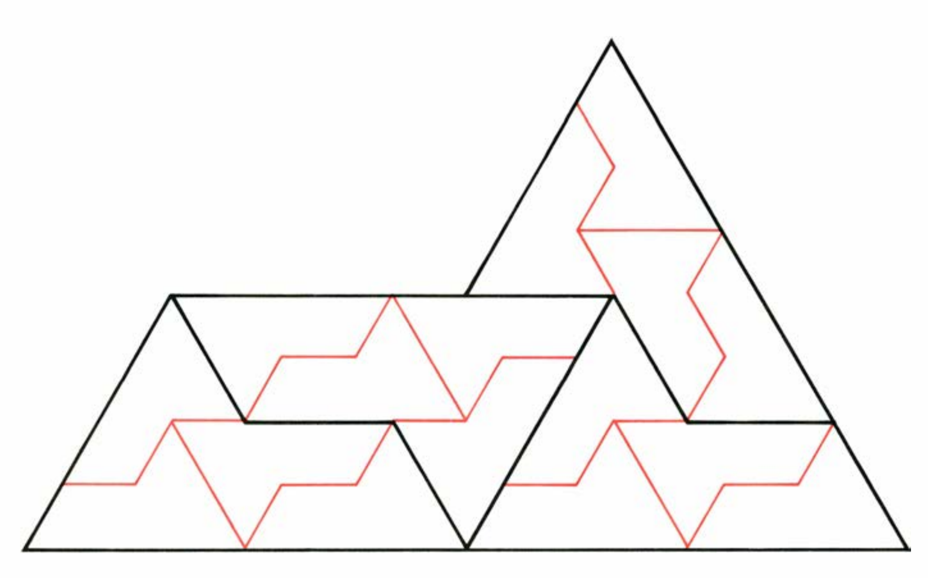


Un eneágono (**en color, izquierda**) *y un par de eneágonos* (**derecha**) *formando un octógono que embaldosa periódicamente*

Mosaico en espiral construido por Heinz Voderberg



Las seis piezas de Raphael M. Robinson que obligan a formar un mosaico no periódico



Tres generaciones de esfinges en un mosaico no periódico

finges. Asimismo, dos esfinges (con un giro de una de ellas de 180 grados) embaldosa periódicamente.

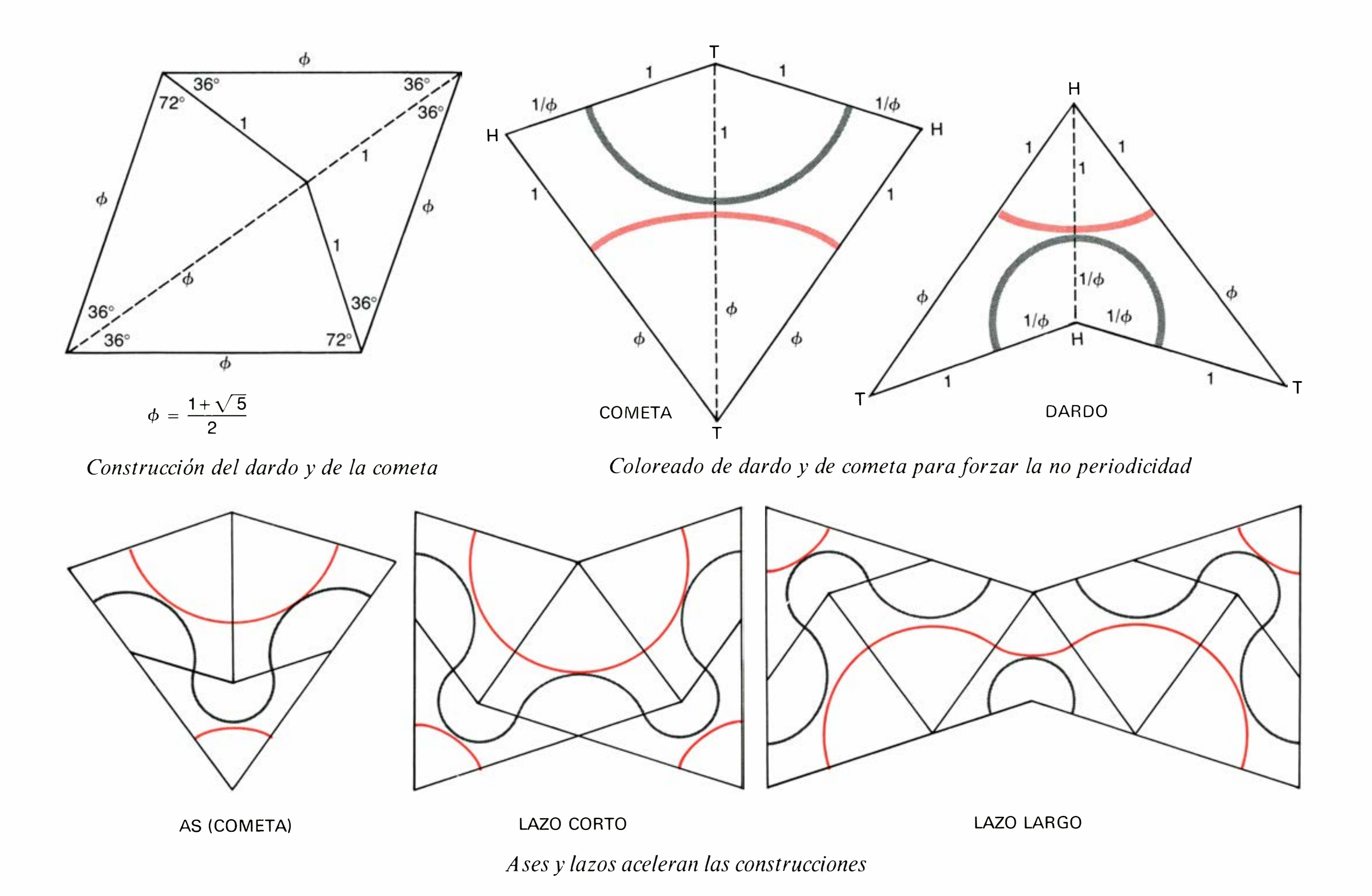
¿Hay conjuntos de mosaicos de dos o más formas *distintas* que embaldosen sólo no periódicamente? Por "sólo" queremos decir que, para construirlo, no puede utilizarse ni una baldosa o subconjunto de ellas ni el conjunto entero de baldosas, periódicamente. Se permite el giro y la reflexión o simetría de las baldosas.

Durante muchas décadas se creyó que tal conjunto no existía, pero se ha demostrado que ello no es así. En 1961, Hao Wang quiso embaldosar el plano mediante conjuntos de cuadrados unitarios, cuyas aristas estuvieran pintadas de varias maneras. Es lo que se llama dominós de Wang. Este pretendía hallar un procedimiento para decidir si un conjunto dado de dominós podrían dar embaldosados si se los colocaba de manera que las aristas contiguas fueran del mismo color, y sin que permitiesen giros ni reflexiones. El problema es importante porque remite a cuestiones de decisión de lógica simbólica. Wang demostró que, si existe, y sólo si existe un procedimiento de decisión, entonces cualquier conjunto de dominós que embaldose el plano no periódicamente, lo embaldosará periódicamente. Wang llegó a la deducción de que un procedimiento tal sí existía.

En 1964, Robert Berger, en su tesis doctoral, defendida en la Universidad de Harvard, sobre matemática aplicada, demostró que la conjetura de Wang era falsa.

No hay un procedimiento general. Por tanto, existe un conjunto de dominós de Wang que embaldosa sólo no periódicamente; Berger construyó tal conjunto usando más de 20.000 dominós. Más tarde, él mismo encontró un conjunto mucho menor, con 104. El año pasado Raphael M. Robinson redujo el conjunto a 24.

Es fácil cambiar un conjunto de dominós de Wang por baldosas poligonales que embaldosen sólo no periódicamente; basta con añadir prolongaciones y ranuras a las aristas, dejándolas en forma de dientes de sierra que encajen en la forma anteriormente indicada, es decir, por colores. Una arista que era anteriormente de un color encaja con otra que era previamente del mismo color, y una relación similar se obtiene para los demás colores. Por medio de giros y simetrías, a partir de dicho mosaico, Robinson construyó seis mosaicos *(véase la ilustración superior izquierda)* que "fuer-



Construcción del dardo y de la cometa

Coloreado de dardo y de cometa para forzar la no periodicidad

Ases y lazos aceleran las construcciones

zan la no periodicidad" en el sentido explicado.

En Oxford, donde enseña matemáticas, Penrose ha venido indagando conjuntos más pequeños. Aunque la mayoría de sus trabajos tratan sobre teoría de la relatividad y mecánica cuántica, siente un poderoso interés por la matemática recreativa, que ya compartía con su padre, el desaparecido genetista L. S. Penrose (fueron los inventores de la famosa "escalera Penrose", que va dando vueltas sin aumentar de altura; Escher la representó en su litografía "Ascensión y descensión"). Si en 1973 Penrose había hallado un conjunto de seis piezas de mosaico que fuerzan la no periodicidad, pronto descubriría una forma de reducirlas a cuatro y, en 1974, las redujo ya a dos.

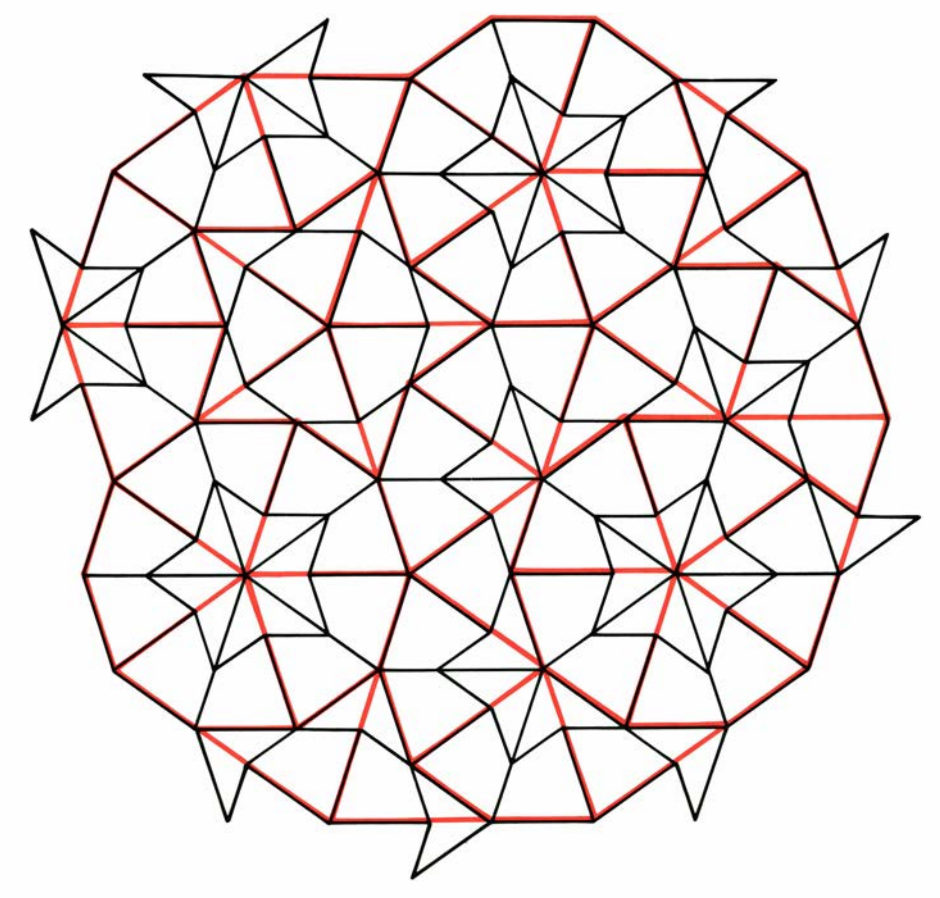
Puesto que los mosaicos se prestan a utilizarse comercialmente, Penrose no quiso revelarlos hasta que no hubo solicitado su patente en Gran Bretaña, Estados Unidos y Japón. Una vez en trámite tal patente obtuve su permiso para escribir a propósito de tales mosaicos. Estoy igualmente en deuda con John Horton Conway por muchos de los resultados de su estudio sobre los mosaicos de Penrose.

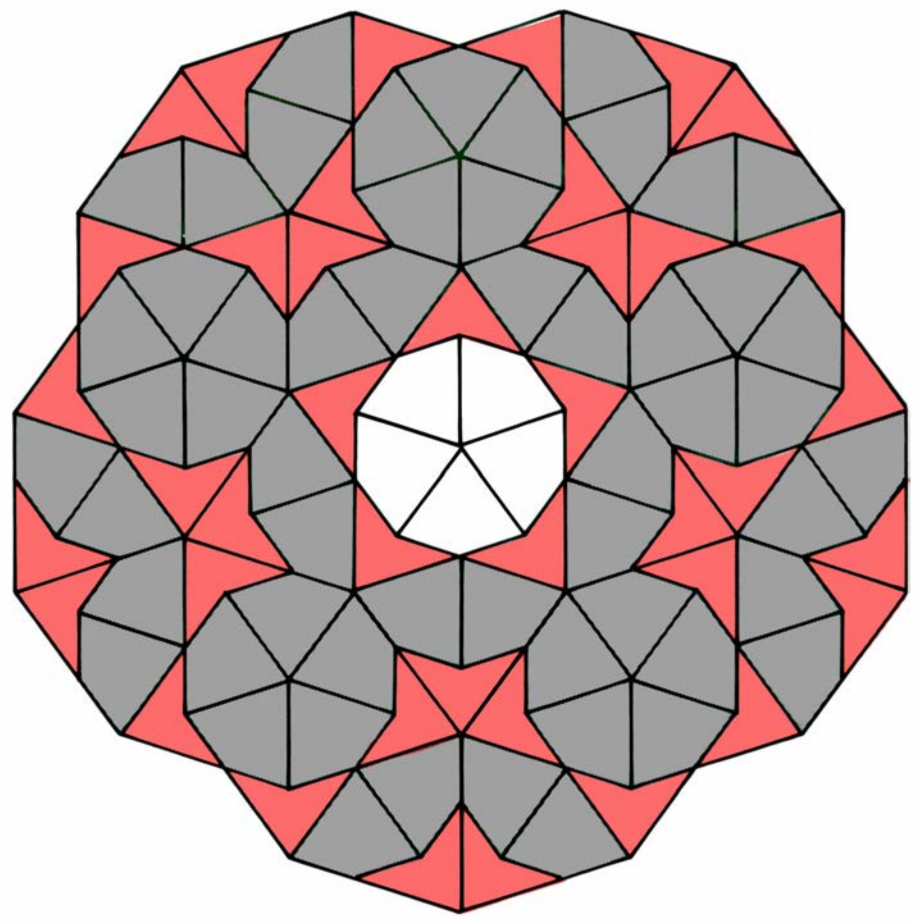
La forma de un par de mosaicos de Penrose puede variar; las del par más interesante son las formas llamadas por Conway "dardos" y "cometas". La ilustración superior izquierda de esta página muestra como derivan de un rombo de ángulos de 72 y 108 grados, por división de la diagonal larga según la conocida razón áurea $(1+\sqrt{5})/2 = 1{,}61803398...$, y uniendo finalmente el punto a los vértices correspondientes a los ángulos obtusos. Sea phi la razón áurea. Cada segmento vale 1 o phi, según queda indicado.

El rombo embaldosa de manera periódica, pero no podemos juntar las piezas de ese modo. Al estar prohibido unir lados de igual longitud, se impone hacerlo con entrantes y salientes, pero hay caminos más simples. Por ejemplo, podemos designar los vértices por H y T, como se muestra en la ilustración superior derecha de esta página; y luego establecer la regla de que sólo pueden unirse las aristas de suerte que coincidan los vértices de la misma letra. En los vértices pueden colocarse puntos de dos colores distintos, para ayudar a que esta regla se cumpla. Otro método más elegante, propuesto por Conway, consiste en dibujar arcos de colores en cada baldosa, según aparece en la ilustración. Cada arco corta los lados y los ejes de simetría, según la razón áurea. Para nosotros, las aristas contiguas deben unir arcos del mismo color.

Si se quiere apreciar bien la belleza y el hechizo del mosaico de Penrose, hay que construir, al menos, 100 cometas y 60 dardos. Las piezas se pintan solamente por un lado. Las áreas de las dos formas están en razón áurea. Esta razón se aplica también al número de piezas necesarias de cada tipo. Se podría pensar que se necesitan más dardos, que son más pequeños; pero no es así. Se necesitan 1,618... veces más cometas que dardos. En un mosaico infinito esta proporción es exacta.

Un camino ideal para construirlo es dibujar, en una hoja, los dardos y cometas que se pueda, pero teniendo en cuenta la proporción de unos 5 cometas por cada 3 dardos y usando una línea fina para las curvas. La hoja puede fotocopiarse muchas veces. Las curvas pueden pintarse con plumas de punta de fieltro rojas y verdes. Conway observó que esto aceleraba las construcciones y mantenía los modelos más estables, si se hacían muchas copias de las tres formas mayores de la ilustración inferior de esta página. Al extender el modelo se puede sustituir siempre dardos y cometas por ases y lazos. De hecho, infinitos *pares,*


Forma de inflar un modelo


El modelo de sol infinito

arbitrariamente grandes, de formas hechas de dardos y cometas servirán para construir cualquier modelo infinito.

Un modelo de Penrose se hace empezando con dardos y cometas alrededor de un vértice, extendiéndolos radialmente. Siempre que se añade una pieza a una arista se tiene que escoger entre un dardo y una cometa de manera que, unas veces, la elección es obligada pero otras no; a veces, las piezas encajan, pero más tarde puede presentarse una contradicción (es decir, un área en donde, legalmente, no puede encajar ninguna pieza), y entonces es forzoso volver atrás y escoger las piezas de otra manera. Una buena actuación consiste en ir colocando, en primer lugar y alrededor del límite exterior, todas las piezas forzosas, las cuales no pueden llevar a ninguna contradicción. Luego se experimenta con las piezas no forzosas. Siempre se puede continuar indefinidamente. Cuanto más se juegue con las piezas, más reglas se aprenden de las que fuerzan una elección, lo que aumenta la eficiencia; por ejemplo, un dardo obliga a poner dos cometas en su concavidad, creando el as ubicuo.

Hay muchas maneras de demostrar que el número de mosaicos de Penrose es no numerable, como ocurre con los puntos de una recta. Estas demostraciones se basan en un fenómeno sorprendente, descubierto por Penrose, al que Conway denomina "inflación" y "deflación". La ilustración superior de esta misma página muestra el comienzo de la inflación: si imaginamos que cada dardo se parte por la mitad, y que las aristas cortas de las piezas originales se pegan entre sí, el resultado es un nuevo mosaico (mostrado en color) de dardos y cometas mayores.

La inflación puede continuarse hasta el infinito, siendo cada nueva "generación" de piezas mayor que la anterior. Nótese que la cometa de la segunda generación, aunque es de la misma talla y forma que el as de la primera generación, se forma de un modo distinto. Por cuya razón al as se le llama también cometa loca. No debe nunca confundirse con otra cometa de la segunda generación. La deflación es el mismo proceso seguido de manera que en cada mosaico de Penrose puedan dibujarse generaciones de dardos y cometas cada vez más pequeños. Modelo que también puede continuarse hasta el infinito.

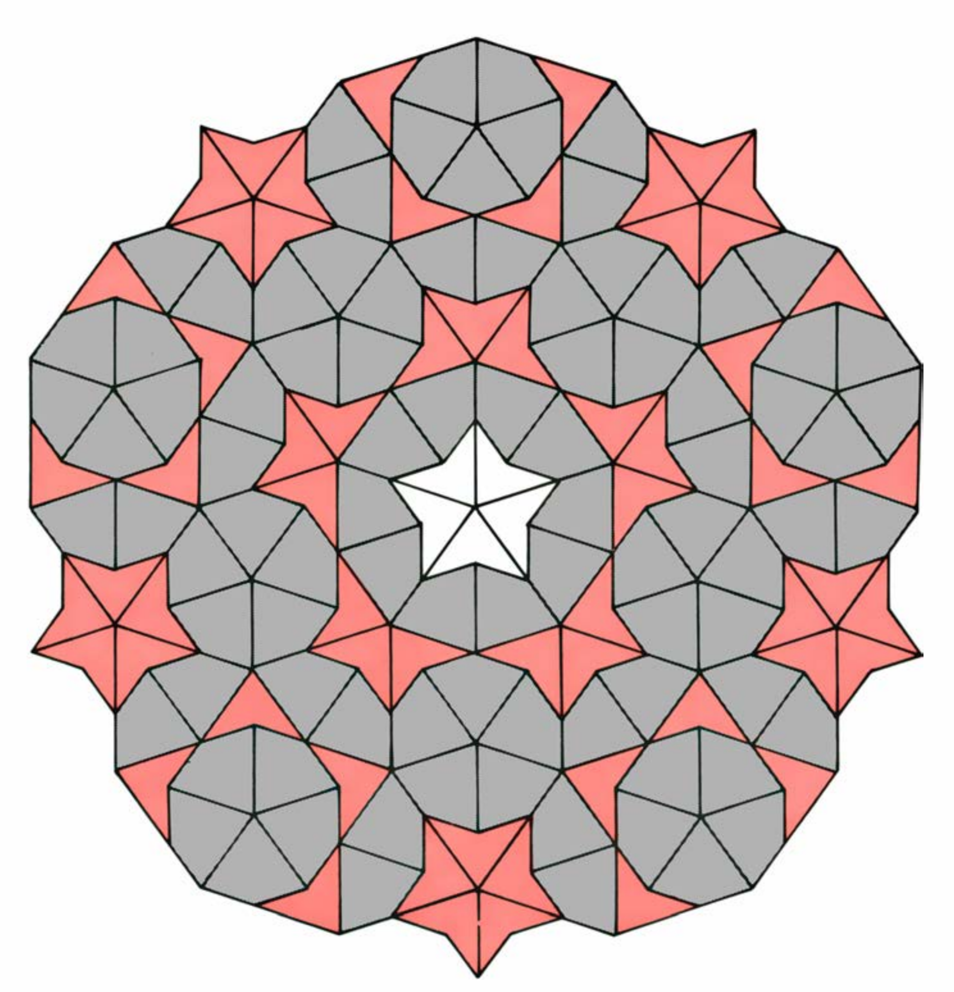
La demostración de Conway de la no-numerabilidad de los modelos de Penrose (quien lo había demostrado antes de un modo distinto) puede esbozarse como

sigue. En una cometa se marca con una *L* un lado del eje de simetría y con *R* el otro, y en un dardo se hace lo mismo usando *l* y *r*. Elegido un punto del mosaico al azar y anotada la letra que da su localización en el mosaico, se aumenta el modelo un paso, y se observa la localización del mismo punto en el mosaico que resulta en la segunda generación anotándose de nuevo la letra. Continuando el proceso de inflación, se genera una secuencia de infinitos símbolos, que es una rotulación única para el modelo original visto, por decirlo así, desde el punto elegido.

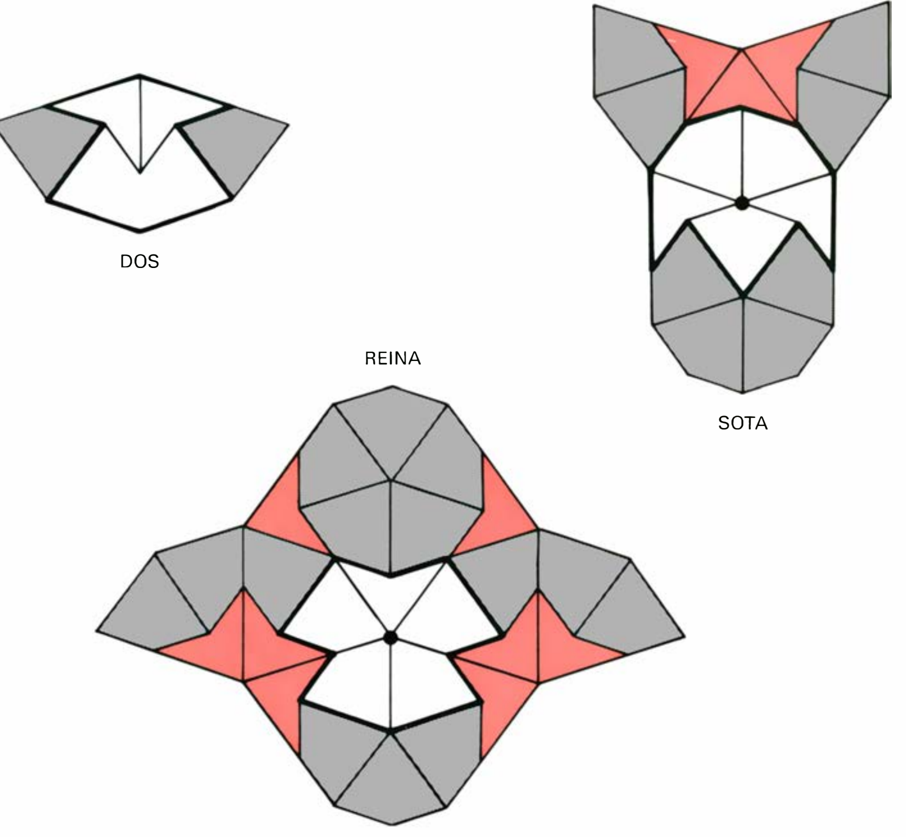
Escogido otro punto del modelo original, el procedimiento puede dar una sucesión que empiece de un modo distinto, pero llegará a formar una letra a partir de la cual coincida hasta el infinito con la primera sucesión. Si no hay tal acuerdo a partir de cierto punto, las dos secuencias rotulan patrones distintos. No todas las posibles sucesiones de los cuatro símbolos pueden producirse de esta manera, pero respecto de las que representan patrones diferentes puede demostrarse que su número se corresponde con el número de puntos de una recta.

Hemos omitido las curvas coloreadas en nuestras figuras de los mosaicos, porque dificultan la observación de las baldosas. Sin embargo, si se trabaja con baldosas de colores, se obtienen un dibujo de extraordinaria belleza. Penrose y Conway han probado, independientemente, que siempre que se cierra una curva, tiene simetría pentagonal y toda la región dentro de la curva tiene simetría pentagonal. En un modelo, puede haber, a lo sumo, dos curvas que no se cierren, pero en la mayoría de modelos las curvas se cierran.

Aunque se puede construir modelos de Penrose con un alto grado de simetría (son múltiples los modelos que tienen simetría bilateral) la mayoría de ellos, como ocurre con el universo, son una engañosa mezcla de orden y de inesperadas desviaciones del mismo. Al extenderse los modelos, parecen esforzarse en repetirse, pero nunca lo logran. G. K. Chesterton llegó a sugerir que un ser extraterrestre, al observar los rasgos del cuerpo humano que se duplican a la izquierda y a la derecha, deduciría razonablemente que tenemos un corazón a cada lado. "El mundo", dijo, "parece un poco más regular y matemático de lo que es; su exactitud es obvia, pero su inexactitud está escondida, su locura está al acecho". En todas partes hay un "silencioso desvío de una pulgada respecto de la exactitud,



El modelo estrella infinito



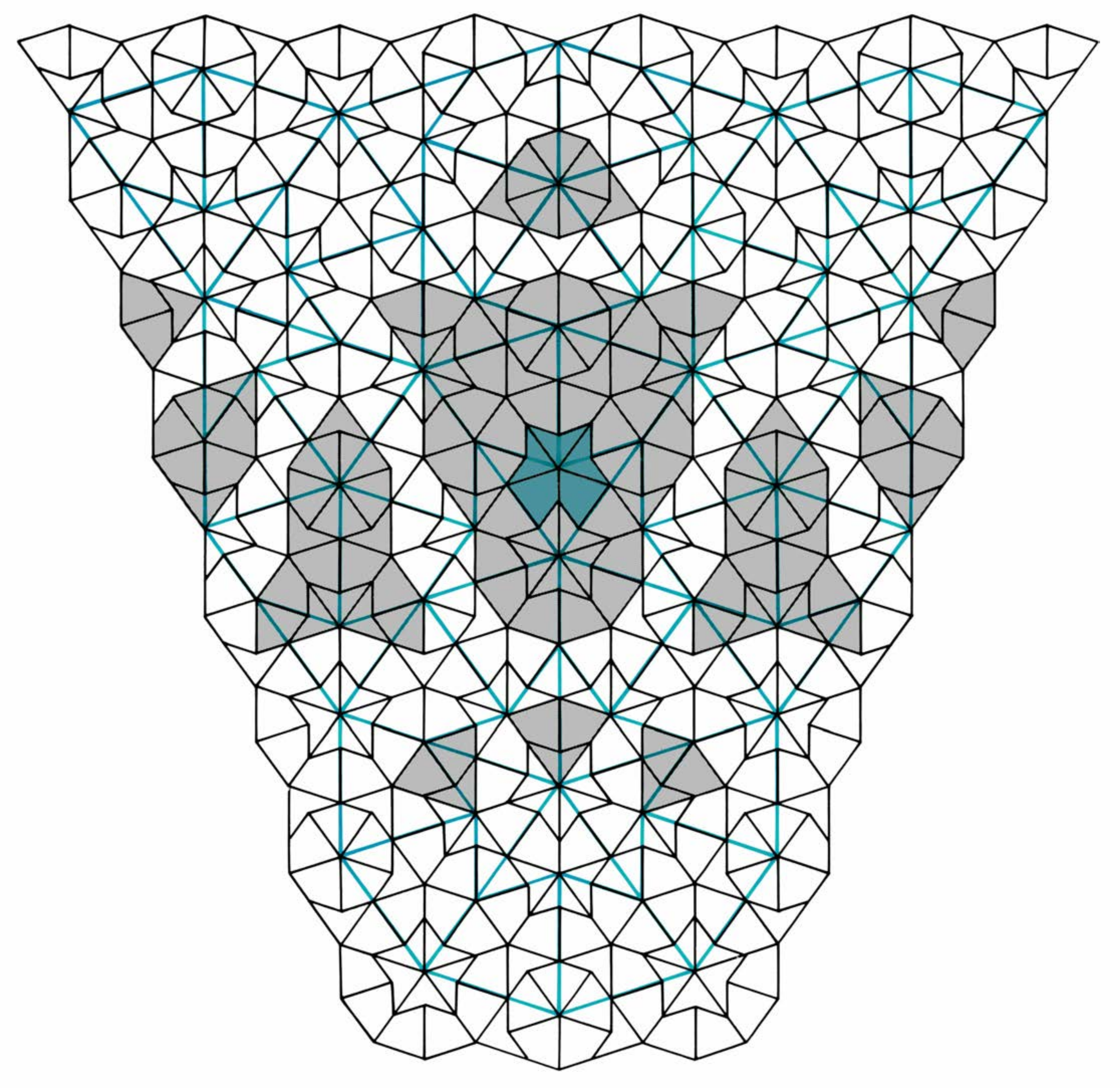
Los "imperios" del dos, la sota y la reina

apenas apreciable, que es un elemento misterioso en todas las cosas... una especie de secreta traición del universo". El pasaje es una bonita descripción de los mundos planos de Penrose.

Hay algo todavía más sorprendente sobre los universos de Penrose. En una curiosa dirección finita, dada por el "teorema del isomorfismo local", todos los modelos de Penrose se parecen. Penrose demostró que cada región finita, de cada modelo, se contiene en alguna parte dentro de cualquier otro modelo. Además, aparece infinitas veces en cada modelo.

Para entender lo singular de esta situación, hay que suponer que se está viviendo en un plano infinito en el que hay un mosaico de Penrose, de entre la infinitud no numerable de los que existen. Dicho modelo puede examinarse pieza a pieza, y por cualquier región. Por largo que sea el tiempo que se esté explorando, nunca puede determinarse en qué mosaico se está, sin que alivie algo ir más lejos y examinar región por región, ya que todas las regiones pertenecen a una gran región finita que se duplica infinitas veces en todos los modelos. Esto es obviamente cierto para cualquier mosaico periódico, pero los universos de Penrose son no periódicos. Se diferencian uno de otro en infinitas maneras, y, no obstante, sólo puede distinguirse uno de otro en el límite inobtenible.

Supongamos que se ha explorado una región circular de diámetro *d*, que consideraremos la "ciudad" en donde vivimos, y que, de repente, nos transportan a un mundo de Penrose paralelo, elegido al azar, ¿en cuánto nos alejamos de una región que tenga exactamente las mismas calles de nuestra ciudad? Conway responde a ello con un teorema notable: ¡la distancia no es nunca superior a 2*d*! (Se trata de un límite superior, no de un promedio.) Si se camina en la dirección adecuada, no se necesita ir más allá de la distancia



El imperio del rey

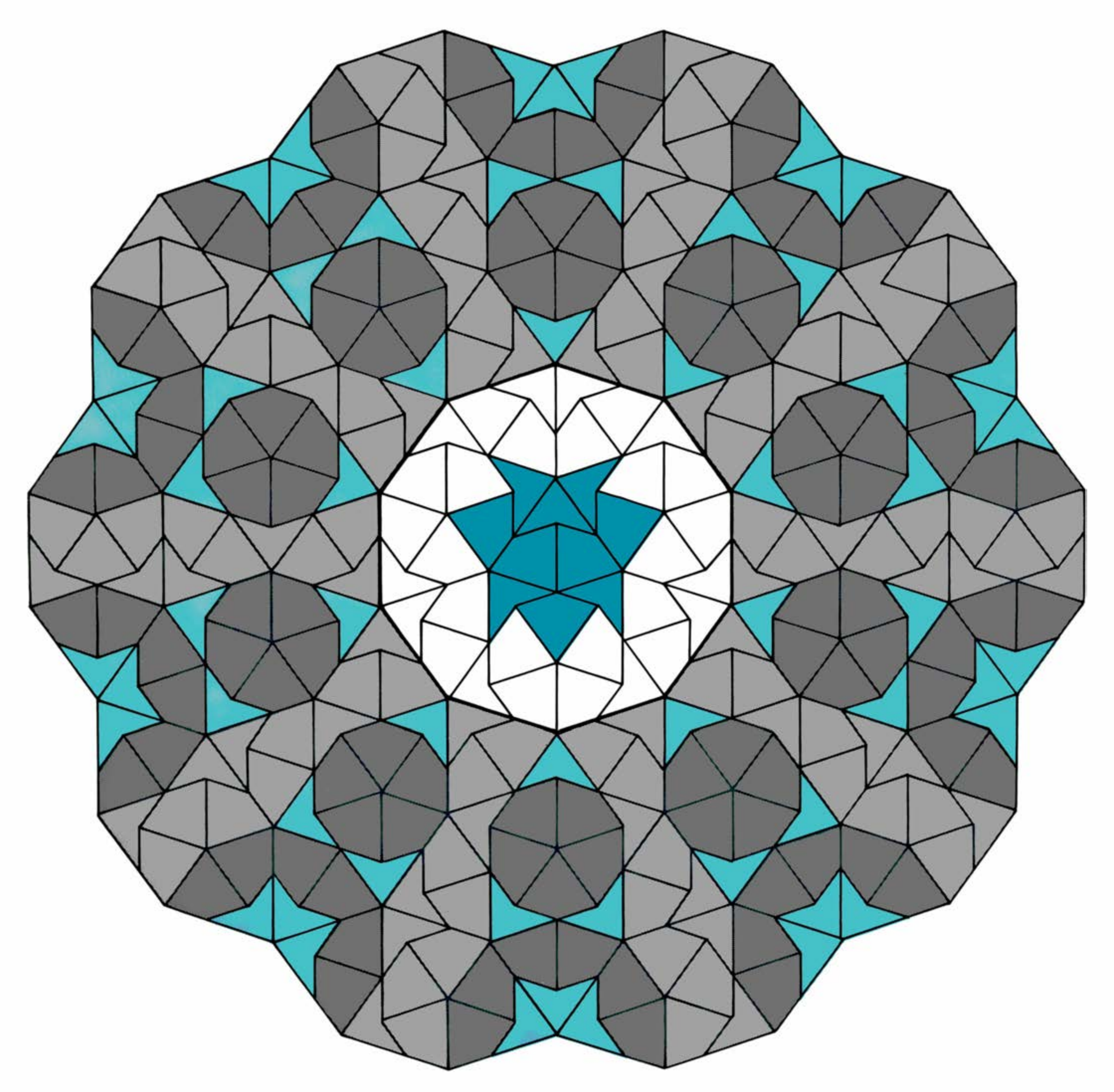
2*d* para encontrarse dentro de una copia exacta de la propia ciudad. El teorema se aplica igualmente al universo en el que se vive. Cada modelo circular grande (hay múltiples de ellos que son distintos) puede alcanzarse andando, en cierta dirección, una distancia inferior a dos veces el diámetro del modelo, y mide, probablemente, su diámetro.

El teorema no podía barruntarse. Considérese un isomorfismo análogo, presentado por una sucesión de dígitos que no siguen ningún modelo, tales como pi. Si se escoge una sucesión finita de 10 dígitos, y se empieza desde un punto al azar en pi, se encontrará ciertamente la misma sucesión, de desplazarse bastante a lo largo de pi, pero a la distancia a la que se tiene que ir no se le conoce cota superior, y la distancia esperada es enormemente superior a 10 dígitos. Cuanto más larga sea la sucesión finita, tanto más lejos se la encontrará de nuevo. En el modelo de Penrose, se está siempre muy cerca de hallar un duplicado del propio hogar.

Hay sólo siete maneras de encajar dardos y cometas en torno de un vértice. Consideremos primero, usando la nomenclatura de Conway, las dos que presentan simetría pentagonal.

El sol (representado en blanco en la ilustración inferior situada en la página 102), no obliga a la colocación de ninguna otra pieza alrededor de él. Si se añaden piezas de manera que se conserve la simetría pentagonal, será forzoso construir el modelo mostrado allí. Está determinado unívocamente hasta el infinito.

La estrella en blanco en la ilustración superior de la derecha (p. 103), obliga a construirse las 10 cometas grises en su alrededor. Al aumentar este modelo, preservando la simetría pentagonal, se creará otro dibujo floreado, infinito y único. Los modelos de estrella y sol son los únicos universos de Penrose

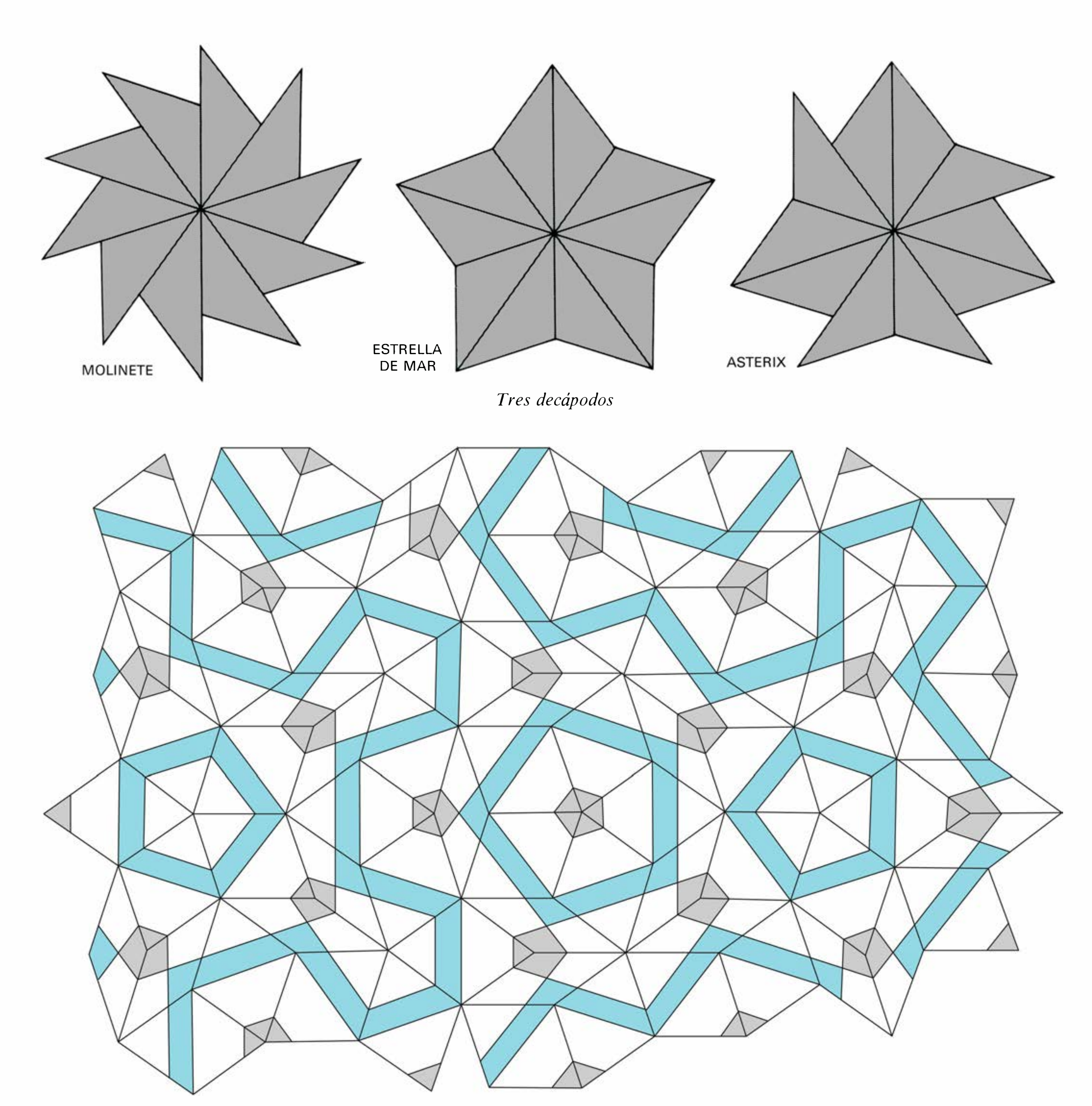


El modelo del carretón

que poseen simetría pentagonal perfecta, y ambos son equivalentes en cierto sentido. Inflando o desinflando cualquier modelo, se obtiene el otro.

El as constituye un tercer modo de colocar baldosas en torno a un vértice, y no obliga a poner más piezas. El dos, la sota y la reina se muestran en blanco en la ilustración inferior de la página 103, rodeados por las baldosas que obligan a poner inmediatamente. Como Penrose descubrió (y más tarde, independientemente, Clive Bach), algunas de las siete figuras del vértice inducen la colocación de baldosas que no están unidas a la región previamente obligada. La ilustración de la pág. 104 muestra en gris lo que quizá sea la mayor parte del "imperio" del rey (el rey es el área coloreada); todas las baldosas grises están condicionadas por el rey. (Hay también dos ases, junto a los límites izquierdo y derecho; son también forzosos, pero no se muestran.)

Este figura del "imperio" del rey fue dibujada por medio de un programa para ordenador escrito por Eric Regener, de la Concordia University, de Montreal. Su programa desinfla todo modelo de Penrose cualquier número de pasos. Las líneas coloreadas muestran el dominio forzado inmediatamente por el rey. Las líneas negras son la tercera generación de desinflamiento, en la que el rey y casi todo su imperio están reproducidos. Se ignora hasta dónde se extiende el imperio, y los imperios menores de la sota y de la reina no han sido trazados en su totalidad.



Tres decápodos

Un mosaico no periódico con los rombos de Penrose

El más extraordinario de todos los universos de Penrose, esencial para entender los mosaicos, es el modelo del carretón infinito, cuyo centro aparece en la página 105 y en la portada. El decágono regular del centro, delineado en trazos negros gruesos (cada lado es un par de aristas, larga y corta), es lo que Conway llamó un "carretón". Cualquier punto de cualquier modelo está dentro de un carretón exactamente igual a éste. Con una inflación de un paso, se ve que cada punto estará dentro de un carretón mayor. De igual forma, cada punto está dentro de un carretón de cada generación, aunque las ruedas no precisan ser concéntricas.

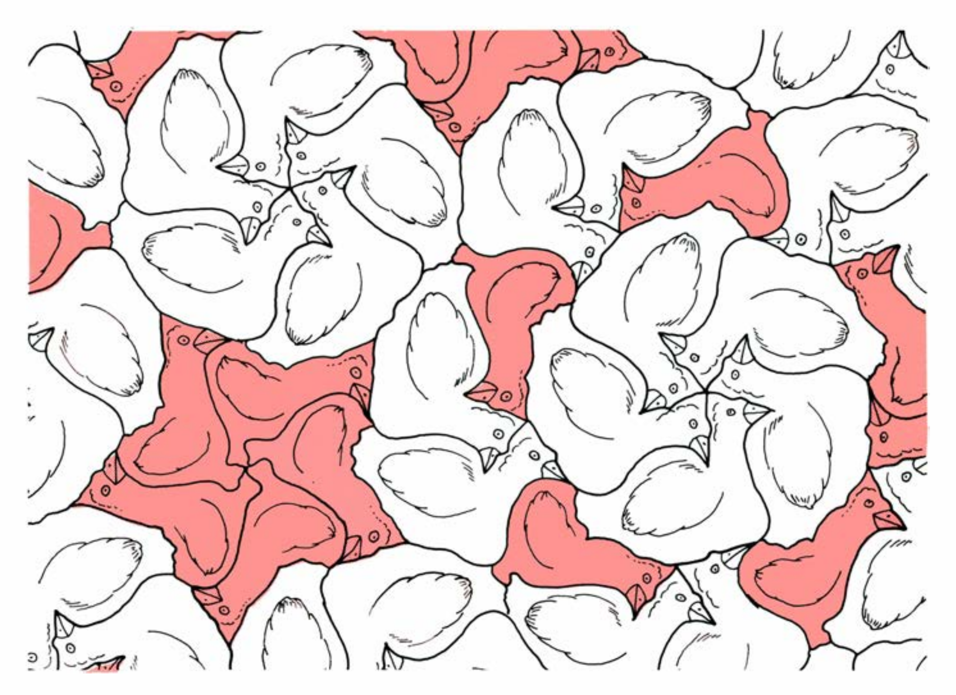
A los diez rayos de rueda grises, que se extienden hacia el infinito, Conway los llamó "gusanos". Constan de lazos cortos y largos; los grandes están en razón áurea con los cortos. Cada universo de Penrose contiene un número infinito de gusanos arbitrariamente largos. Inflese o desínflese un gusano, y se obtiene otro gusano a lo largo del mismo eje. Obsérvese que dos gusanos completos se extienden a través del carretón central en el modelo del carretón infinito (en su interior no hay grises). Los restantes rayos son medios gusanos. Aparte de los rayos y del interior del carretón central, el modelo tiene una simetría decagonal perfecta. Entre dos rayos cualesquiera se ve una disposición alternada de grandes porciones de los modelos del sol y de la estrella.

Los rayos del modelo del carretón infinito pueden moverse de un lado a otro (o lo que es equivalente, cada uno de sus lazos puede girarse extremo contra extremo); y el rayo sigue encajando con todas las baldosas que lo rodean, excepto con aquellas que están dentro del carretón central. Hay 10 rayos, por tanto tendremos $2^{10}=1024$ combinaciones de estados. Sin embargo, tras eliminar giros y simetrías, sólo quedarán 62 combinaciones distintas, cada una de las cuales deja dentro del carretón una región que Conway denominó un "decápodo".

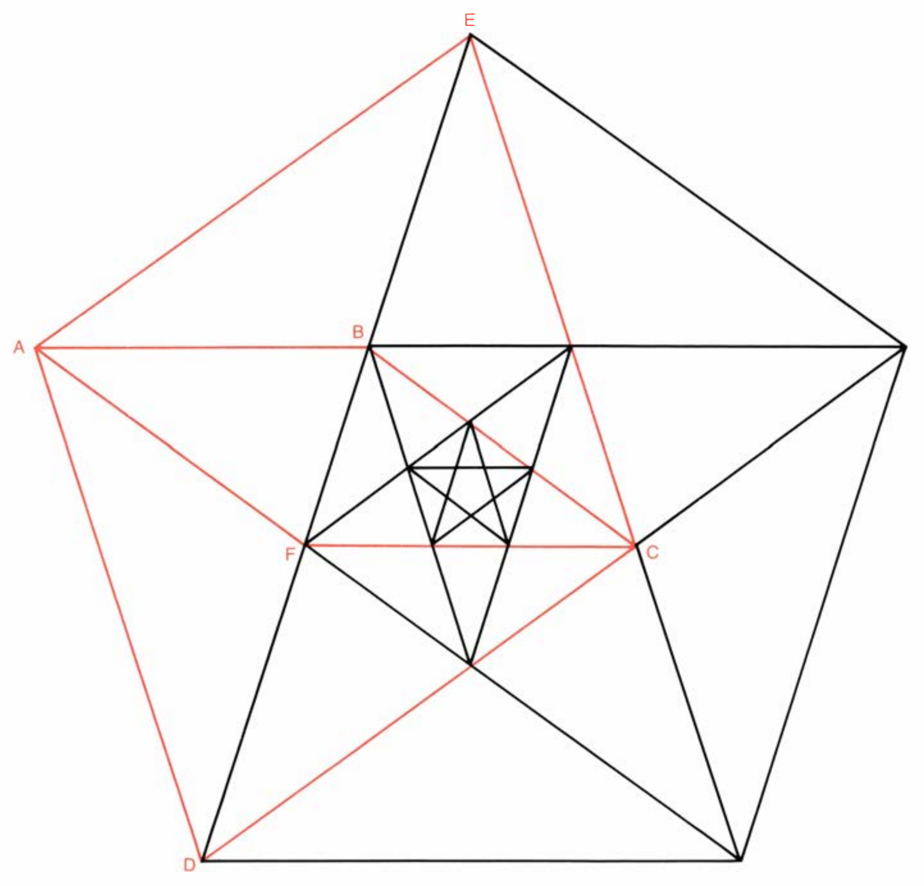
Los decápodos se construyen con 10 triángulos isósceles idénticos, de la forma de medio dardo. Los decápodos de máxima simetría son el "molinete" y la "estrella de mar". Al igual que un gusano, cada triángulo puede moverse como antes. Sin recurrir a giros ni a las simetrías, se obtienen 62 decápodos.

Cuando los rayos están dispuestos tal cual se hallan en el modelo del carretón infinito, en el centro se forma un decápodo llamado Batman. Este (en color oscuro) es el único decápodo que puede ser embaldosado legalmente. (Ninguna región finita puede tener más de un mosaico legal.) Sin embargo, Batman no obliga a formar el modelo del carretón infinito, simplemente lo permite. De hecho, ninguna porción finita de un mosaico legal puede forzar un modelo entero, ya que la porción finita está contenida en *cada* mosaico.

Nótese que el modelo del carretón infinito es simétrico bilateralmente; sus ejes de simetría pasan verticalmente a través de Batman. Inflado el modelo,



Las gallinas no periódicas de Penrose



El pentagrama pitagórico

se mantiene invariable, excepto por simetría especular, en una línea perpendicular a los ejes de simetría. Los cinco dardos Batman y sus dos cometas centrales son las únicas baldosas en todos los universos de Penrose que no están dentro de una región con simetría pentagonal. Todas las demás piezas, en este modelo o en cualquier otro, están en infinitas regiones con perfecta simetría pentagonal.

Los otros 61 decápodos se producen dentro del carretón central mediante las

Los otros 61 decápodos se producen dentro del carretón central mediante las otras 61 combinaciones de giros de gusanos de los rayos. Todos son "agujeros" en el sentido siguiente: un agujero es una región finita vacía, rodeada por un mosaico infinito, que legalmente no puede embaldosarse. Podría suponerse que cada decápodo es el centro de infinitos mosaicos; pero aquí los universos de Penrose nos gastan otra broma. De modo sorprendente, 60 decápodos obligan a formar un único mosaico que difiere del expuesto sólo en la composición de los rayos. Batman y un decápodo llamado Asterix, en memoria del personaje francés de las historietas infantiles, no lo hacen. Asterix permite un modelo de carretón infinito, pero también modelos de tipos distintos.

Y ahora una asombrosa conjetura: Conway cree, aunque no lo ha demostrado plenamente, que cada agujero, de cualquier tamaño o talla, equivale a un decápodo; y ello, en el sentido de que, al reorganizar las piezas de mosaico alrededor del agujero y sacar o añadir un número finito de piezas si es necesario, puede transformarse un agujero en un decápodo.

Si ello es verdad, cualquier número finito de agujeros en un modelo puede reducirse a un decápodo. Sólo hay que mover un número suficiente de baldosas para unir los agujeros en otro nuevo mayor, y entonces reducir este último hasta que resulte un decápodo no embaldosable.

Imaginemos que un decápodo es una pieza de mosaico sólida. A excepción de Batman y Asterix, cada uno de los 62 decápodo es como una imperfección que solidifica un cristal. Induce un modelo único de carretón infinito, rayos incluidos, que prosigue continuamente. Si la conjetura de Conway se cumple, cualquier "pieza extraña" (término usado por Penrose) que obliga a formar un mosaico único tiene un contorno que se transforma en uno de los 60 decápodos, por grande que pueda ser la pieza.

Las cometas y los dardos pueden cambiarse en otras formas por la misma técnica descrita anteriormente, cambiando triángulos isósceles en mosaicos espirales de polígonos. Es la misma técnica que Escher empleó para transformar piezas de mosaico poligonales en formas animales.

La ilustración superior de la página precedente muestra como Penrose cambió sus dardos y cometas en gallinas que embaldosan de manera no periódica. Notése que aunque las gallinas son asimétricas, nunca hay que darles la vuelta a ninguna de ellas para disponer el mosaico en el plano. Desgraciadamente, Escher murió antes de conocer los mosaicos de Penrose. ¡Cómo habría gozado con sus posibilidades!

Disecando los dardos y cometas en piezas menores y juntándolas de otras maneras, pueden hacerse nuevos pares de piezas de mosaico de propiedades similares a los dardos y cometas. Penrose encontró un par simple y poco corriente: los dos rombos de la ilustración inferior de la página 106. Todas las aristas poseen la misma longitud. La pieza más larga tiene ángulos que miden 72 y 108 grados y, las más pequeñas, ángulos de 36 y 144 grados. Como antes, las áreas y el número de piezas necesarias de cada tipo están en la razón áurea. Los modelos de mosaicos pueden inflarse y desinflarse, y embaldosan el plano en una infinidad no numerable de formas no periódicas distintas. La no periodicidad puede inducirse mediante entrantes y salientes o coloreando, según sugiere Penrose y se muestra en la ilustración.

En el pentagrama de la ilustración inferior de la página 107, se ve la estrecha proximidad de los dos conjuntos de piezas de mosaico, uno respecto del otro y con relación a la razón áurea. Se trata del símbolo místico de la secta pitagórica y el diagrama con el que Fausto atrapó a Mefistófeles. La construcción puede continuar indefinidamente hacia dentro y hacia fuera, y cada segmento está en razón áurea con el siguiente más pequeño. Nótese que las cuatro baldosas de Penrose están incluidas en el diagrama.

La cometa es $ABCD$ y el dardo es $AECB$. Aunque los rombos no están en las justas proporciones relativas, son $AECD$ y $ABCF$.

¿Existen pares de baldosas no basadas en la razón áurea que obliguen a formar mosaicos no periódicos? ¿Existe una única pieza que embaldose sólo de modo no periódico? Estas preguntas definen dos de los problemas más intrigantes que permanecen sin resolverse en la teoría de los mosaicos.

Taller y laboratorio

Construcción de un microscopio simple de Anton Leeuwenhoek

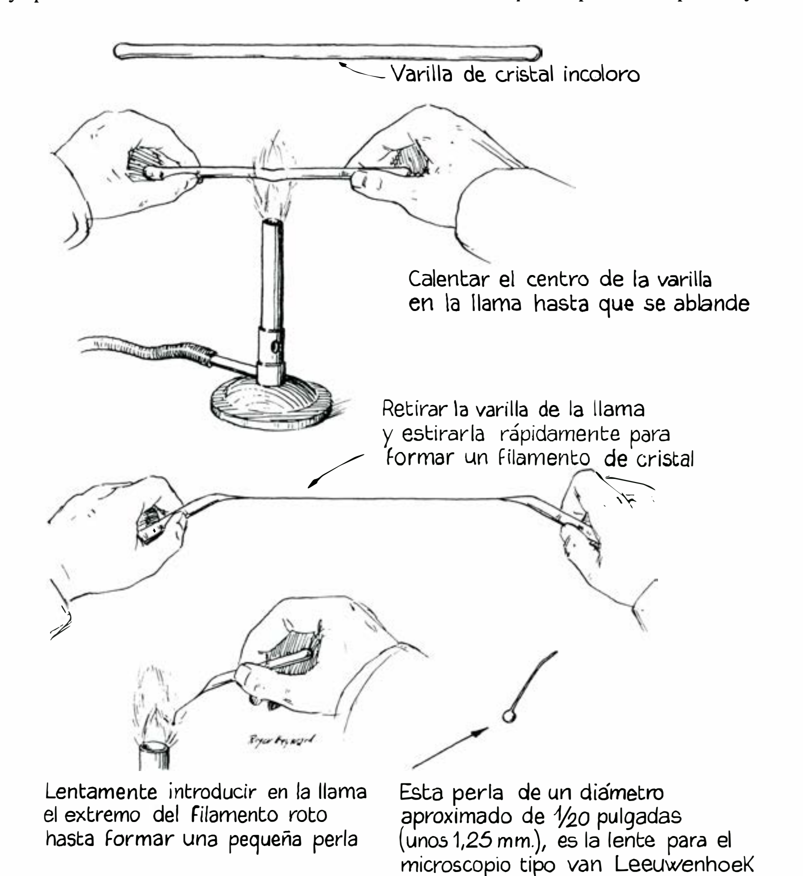
C. L. Stong

La falta de visión de Adán, cuando dio nombre a las criaturas de la tierra (*Génesis,* 2:19), puso las cosas realmente difíciles para sus descendientes científicos. Si hubiera hecho una lista de los animales según los nombraba, ¡cuán fácil sería ahora, por ejemplo, clasificar una preparación para microscopía! Tal como están las cosas, el redescubrimiento y nueva denominación de los organismos del mundo ha supuesto una lenta y penosa labor. Aristóteles conocía unos 520 animales y Teofrasto podía identificar aproximadamente otras tantas plantas. Hoy en día, gracias principalmente a Linneo y a la invención del microscopio, nuestro catálogo ha aumentado hasta cerca de un millón de especies animales y 336.000 especies de plantas. Sin embargo, el censo de la vida sobre la tierra está lejos de ser completo; nadie sabe cuántos miles de especies quedan por descubrir y clasificar.

La búsqueda para completar y ordenar la colección de organismos constituye uno de los desafíos más agradecidos de la ciencia. Y para los aficionados a ésta, se trata de un deporte apasionante. Se puede participar con solo adquirir un microscopio, ya que la zona más extensa por explorar está ocupada por los microorganismos.

La microscopía presenta para el aficionado otros atractivos además del hallazgo de nuevos organismos.

Puede construirse en casa un microscopio útil, en menos de una hora, sin tener que desplazarse mucho para encontrar material de estudio. La saliva, por ejemplo, proporciona suficiente variedad de muestras a estudiar durante meses. Parece como si cuantos más organismos hallamos e identificamos, más quedan por descubrir. El problema de la miscroscopía no es tanto encontrar materia de estudio cuanto desarrollar la voluntad de querer dedicarse a una cosa concreta. Constantemente nos vemos tentados a explorar los nuevos campos que nos abre el instrumento. Supongamos que alguien derrama la sal y que sus cristales le llaman la atención. Uno empieza a pensar en la cristalografía. Basta con ir hasta la cocina para encontrar material más que suficiente para mantener su microscopio ocupado durante horas, examinando la morfología cúbica de la sal o la estructura brillante del azúcar. Una partícula de polvo, extraída de la punta de los zapatos, proporciona una amplia colección de minerales cristalinos: cuarzo, mica, sílice, calcita. Fragméntese un cubito de hielo en el vaso, póngase una lámina sobre el portaobjetos y obsérvese rápidamente. Los cristales aciculares se convierten casi instantáneamente en una esfera reluciente. Comprobar la pureza de la gota de agua resultante. ¿Contiene



Construcción de una lente de perla de un microscopio de Leeuwenhoek

partículas de materia en suspensión o quizá organismos en estado latente? Examinar unos pocos granos de pimienta, ¿ha sido adulterada, como a veces ocurre, mediante la adición de almidón? Si así es, se identificarán en seguida los granos de forma ovalada. ¿Se está interesado en las grasas? Contrastar el aspecto de una mancha de mantequilla con un poco de grasa de carne cocida. El interés inicial por los cristales puede ampliarse en cuestión de minutos.

Los historiadores no están seguros de quién inventó el microscopio. Como muchos productos de la tecnología, parece ser que el instrumento ha evolucionado a partir de diversos conocimientos acumulados, muy entrelazados y difíciles de aislar.

La lupa más antigua de las descubiertas hasta ahora fue encontrada en las ruinas de Nínive por el arqueólogo británico Sir Austen Layard. Era una lente planoconvexa de cristal de roca toscamente pulida, que aumentaba bastante bien. Plinio el Viejo, en el año 100 p.C., citó la "propiedad de quemar que tenían las lentes hechas de cristal". Pero la ciencia de la óptica, según se entiende modernamente, no se inició hasta el siglo XIII, aproximadamente.

Roger Bacon parece haber sido el primero en sugerir sus principios. Sus escritos serían el fundamento del microscopio y del telescopio, y, probablemente, se le puede otorgar la invención de las gafas.

Bacon enseñó la teoría de las lentes a un amigo, Heinrich Goethals, el cual visitó Florencia en 1285. La información de Goethals pasó a un tal Salvina D'Armato a través de un fraile dominico, Spina. La tumba de D'Armato en la iglesia de Sta. Maria Maggiore lleva la siguiente inscripción: "Aquí yace Salvina D'Armato de los Amati de Florencia, inventor de las gafas. El Señor perdone sus pecados. A.D. 1317".

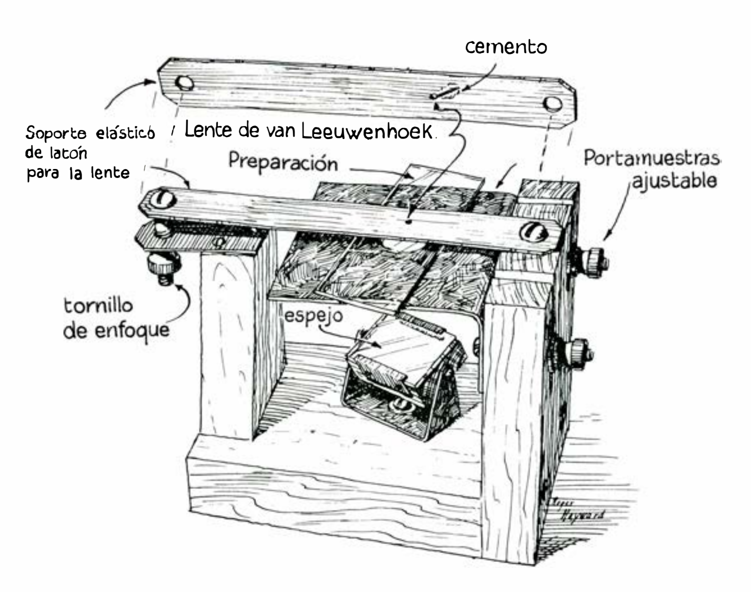
El microscopio simple –de una sola lente– debió usarse tan pronto como se inventaron las gafas, e incluso pudo precederles. No se sabe quien fue el primero en utilizarlo. Sin embargo, el primero que se sirvió de él para algunos descubrimientos importantes fue el holandés Anton van Leeuwenhoek, nacido en 1632. Tras examinar algunos materiales corrientes con un instrumento simple de lente única, que él mismo había construido, escribió emocionado a la Royal Society de Londres acerca de todos los increíbles objetos que revelaba.

Descubrió "organismos culebreantes" y "gusanos" en el agua tomada del canal de su Delft natal y en raspados de sus dientes. Tal vez su contribución más importante fue la observación de los glóbulos rojos de la sangre. Leeuwenhoek no sólo identificó los eritrocitos sino que también realizó dibujos precisos de su forma enviándolos, junto con las mediciones de su tamaño, a la Royal Society.

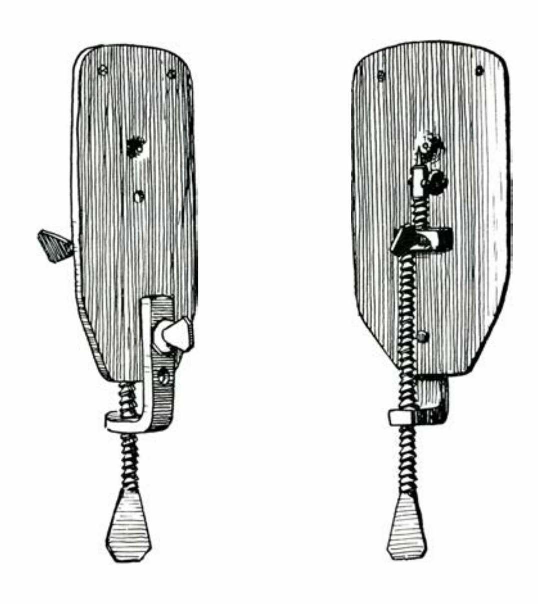
Cualquiera, con un poco de tiempo libre, puede construirse un microscopio de Leeuwenhoek. Su ejecución es fácil y confiere al principiante una valiosa experiencia en la preparación y manejo de los especímenes.

Se necesita una varilla de cristal, corta, y delgada; un trozo de hierro o latón de 2,5 × 7,5 cm y de 1,5 mm de espesor, aproximadamente; dos tornillos para metales, con las correspondientes tuercas; un tubo de cemento de secado rápido y un poco de celofán.

En cuanto a la varilla de cristal, bastará una de vidrio incoloro y transparente, de las utilizadas para agitar líqui-



Microscopio de Leeuwenhoek adaptado a los modernos portaobjetos



Un microscopio de Leeuwenhoek

dos en los laboratorios. Se pasa el centro de la varilla por la llama de un mechero Bunsen o de un fogón de la cocina. La varilla se introduce gradualmente en la llama para evitar las fuerzas creadas al calentarla bruscamente, que causarían su rotura. El centro de la varilla alcanza rápidamente el rojo vivo y se torna maleable.

Retirar rápidamente la varilla del fuego y estirarla. Su centro ablandado dará lugar, al estirar, a un filamento a modo de cabello de unos 60 cm de longitud.

Una vez enfriada la varilla, se rompe un fragmento, de unos quince centímetros, de la parte media del filamento. Paulatinamente se va poniendo en contacto un extremo de este hilo con la llama. Su extremo se volverá incandescente casi en el acto, formándose una pequeña perla. Seguir introduciendo el filamento en la llama hasta que la perla alcance un diámetro aproximado de unos 1,5 mm.

La lente del microscopio ya está terminada. La pequeña perla, si se ha preparado cuidadosamente, tendrá aproximadamente un aumento de 160 diámetros. (El aumento de una lente de este tipo viene a ser igual al número 10 dividido por el diámetro de la perla expresado en fracciones de milímetro.) La calidad de las lentes fabricadas por este método primitivo está lejos de ser uniforme; por consiguiente, deben hacerse varias y seleccionar la mejor.

Unido a la perla, puede dejarse un segmento del filamento de cristal, utilizándolo luego para montar la lente en su soporte. Leeuwenhoek montó su lente entre dos placas de latón con un orificio para aquélla. Pero a mí me ha resultado más conveniente perforar una sola placa y pegar a ella la pieza de cristal por su vástago, con la perla ocluyendo el agujero. El agujero debe ser ligeramente menor que la perla, con el fin de que no se escape la luz por los lados de la lente, cosa que disminuiría el contraste de la imagen. El cristal se fija a la placa con cemento de secado rápido.

La distancia focal de esta lente minúscula es muy corta. Esto significa que el portaobjetos sobre el que se monte el espécimen debe hallarse muy cerca de la lente, a veces casi tocándolo. Para enfocar el microscopio y situar la muestra en la posición correcta, Leeuwenhoek se sirvió de un conjunto de tornillos que movían una punta de metal, la cual servía de portaobjetos. En mi versión de su instrumento, he sustituido la punta metálica por un trozo de celofán y lo he pegado con cemento al mecanismo de ajuste. Las muestras se engloban en el celofán.

Por desgracia, el microscopio de Leeuwenhoek carece de la comodidad de observación de los instrumentos modernos. Para ver la imagen aumentada hay que acercar mucho la lente a la vista. Roger Hayward diseñó un modelo mayor, empleando un portaobjetos clásico como portamuestras, un espejo para controlar la luz y un control del enfoque más práctico. Estas modificaciones mejoran la comodidad del instrumento, pero no eliminan la necesidad de acercar la lente al ojo.

Teniendo en cuenta su primitivo diseño, el microscopio de Leeuwenhoek revela una asombrosa cantidad de detalles. Se supone que Leeuwenhoek trabajó con éxito con perlas más pequeñas, que daban mayores aumentos, pero rápidamente aprendió a valorar más la resolución que el poder magnificador, y a trabajar con el aumento más bajo posible. Una imagen grande, pero borrosa, no ofrece ninguna ventaja sobre otra pequeña y también borrosa.

Leeuwenhoek, dictó, por lo menos, otra lección fundamental, a saber: la importancia de preparar cuidadosamente los objetos para su examen microscópico. Como escribiera, en el siglo XVIII, el matemático Robert Smith en su *Compleat System of Optiks* "tampoco debemos olvidar una habilidad en la que él (Leeuwenhoek) sobresalió muy particularmente, la de preparar sus muestras de la mejor manera para ser observadas con el microscopio; y estoy seguro que cualquiera que examine algunas de estas mismas muestras a través de estas lentes, quedará satisfecho. Por lo menos yo mismo he encontrado mucha dificultad a este respecto, observando una diferencia muy apreciable entre los detalles de la misma muestra preparada por mí y la preparada por el Sr. Leeuwenhoek, observada con lentes de calidad muy parecida".

Desde la época de Smith, generaciones enteras de fabricantes de portaobjetos han desarrollado técnicas para la preparación de muestras, tan fascinantes casi como el manejo del propio microscopio.

Algunas muestras bastante notables, como la raíz seca de un cabello, o la pulga muerta de un perro, no requieren otra preparación que la de colocarlas sobre un portaobjetos con una pizca de bálsamo del Canadá, o algún otro englobante para preparaciones, y cubrirlas con un delgado cristal. Los objetos diminutos, como los glóbulos rojos de la sangre, pueden verse bastante bien, si se extienden con cuidado sobre un portaobjetos y se protegen con un cubreobjetos. Pero los gruesos y opacos, o transparentes, o los que contienen agua en su estructura, requieren un tratamiento especial.

Si hay que estudiar el interior de una muestra, debe cortarse o, si es transparente, hay que iluminarla por debajo. Algunas muestras deben cortarse en secciones muy finas. Hay aparatos para cortar llamados microtomos, que pueden cortar tejido congelado o tejido incluido en cera, en secciones casi tan finas como la longitud de las ondas luminosas. Además de resolver el problema de la iluminación, los cortes finos simplifican la imagen, ya que el microscopio aumenta en todas direcciones. No se necesita ninguna máquina especial para cortar. Una hoja de afeitar sirve para muchos tipos de muestras.

Cuando un organismo es completamente transparente, a veces debe colorearse o colocarse dentro de una substancia refractora de la luz. El proceso de tinción es un arte en sí mismo, pues ésta produce siempre una mutación química en el organismo. Algunos colorantes afectan a una parte de la célula, pero no a otras. Utilizando diferentes productos químicos se puede teñir el núcleo de un color y el citoplasma que lo rodea de otro.

Muchas bacterias pueden distinguirse unas de otras únicamente por la forma en que captan un determinado colorante; ésta es la base, por ejemplo, de la clasificación de las bacterias en "gram positivas" y "gram negativas".

La preparación de muestras para el microscopio ha dado lugar a su propia literatura especializada, con volúmenes enteros dedicados a temas tales como técnicas de desecación; limpieza; blanqueado para extraer los pigmentos que perturban la visión; métodos de flotación de muestras en células líquidas; selección de cubreobjetos con propiedades ópticas que concuerdan con las del instrumento; y el pulido y corrosión de superficies metálicas para revelar la estructura cristalina. Procesos todos ellos que resultan casi tan numerosos y variados como los objetos que pasan enfocados por la lente del observador.

Después de construir y utilizar un aparato de Leeuwenhoek, es muy probable que uno decida ir más allá. Un microscopio compuesto clásico ahorrará mucho esfuerzo visual. Debe ser un instrumento de buena calidad, capaz de mostrar finos detalles. Sus aumentos han de concordar con la capacidad del principiante. Los objetivos de gran aumento suelen resultar defraudantes para aquél, ya que su buena utilización requiere destreza.

Libros

El cuerpo no segmentado de la moscarda es una perfecta máquina creada para comer

Philip Morrison

THE HUNGRY FLY: A PHYSIOLOGICAL STUDY OF THE BEHAVIOR ASSOCIATED WITH FEEDING, por V. G. Dethier. Harvard University Press. El insecto *Phormia regina* es una "joya a chorro" de 30 miligramos, tres esferas imperfectas en línea, alas diáfanas, grandes ojos compuestos, cuerpo encapsulado en armadura iridiscente, semirrígida y tostada, "una perfección en sus partes". El profesor Dethier, biólogo de la Universidad de Massachusetts, convivió durante treinta años con esta especie que tiene 60 millones de años de antigüedad, sometiéndola a numerosas pruebas. Con una veintena de licenciados y muchos colegas de todo el mundo, ha construido un modelo sencillo y vigoroso del comportamiento de esta moscarda en lo que se refiere a su alimentación (se deja aparte la tarea más importante en su vida, la reproducción). Expone viejos y nuevos experimentos; la historia de la experimentación ilumina con gran agudeza las conclusiones y la técnica seguida. Algunas pruebas son de extrema simplicidad, como dejar que la moscarda camine sobre un trozo de papel humedecido con una solución azucarada teñida de azul. Las "huellas labiales" del moscón registran su andadura. Otros corresponden al virtuosismo instrumental de nuestro tiempo: micrografías de barrido electrónico y encefalogramas del insecto. Dethier busca analogías o clarificadoras disimilitudes con otras especies, desde la rata hasta el hombre.

Las moscardas, además de hermosas, son complicadas. Con sus ojos sólo pueden rivalizar los de los cefalópodos y los de los vertebrados; cuando el neurofisiólogo español Santiago Ramón y Cajal "contempló por primera vez los ojos de las moscas, no pudo menos de sentirse asombrado ante su enorme complejidad". Están soberbiamente adaptados para descubrir el movimiento. Intente cazar una mosca. Ven todo el espectro visible, detectan la polarización y pueden distinguir los fotogramas sucedidos a una velocidad 20 veces mayor que nuestra capacidad de resolución temporal. Su diminuto cerebro, dotado de 100.000 neuronas, recoge datos del ojo y señales de los miles de sensores químicos, táctiles y olfatorios instalados en los flagelos y pelillos de esa armadura mosquil, conjuntamente con muchas señales internas.

La primera necesidad que manifiesta la mosca es de comer. Vuela al azar, registrando las corrientes de aire para ir contra ellas hacia la fuente de olor apetecible, mientras vigila las formas visibles para medir su velocidad respecto del suelo. Vuela en círculos concéntricos luego, hasta que alguna señal óptica demuestre la conveniencia de aterrizar. Camina por superficies inertes en cortos paseos. Si la superficie huele mal, la mosca remonta nuevamente el vuelo. Sobre una superficie sabrosa (sus patas son catadoras) sigue un patrón bastante regular. La mosca gira cada vez que una pata pierde contacto con la comida. O bien "danza", es decir, da un paso estereotipado aunque "ruidoso" dentro de una pequeña zona, cuya forma responde a la inclinación de la superficie, al menos en la obscuridad. No cabe duda de que se trata de una premonición de la danza solidaria de la abeja.

Una vez localizado el alimento de tan compleja forma, la mosca bebe. Al carecer de boca, se limita a sorber líquidos a través de su trompa tubular. La mosca debe elegir los alimentos aceptables y rechazar los nocivos o desagradables (los alimentos silvestres no son nunca puros). Sabe cuándo tiene que suspender la comida, como sabe cuándo debe volver a buscarla. Bebe agua dulce; según parece, puede detectar vapor de agua, como también otros compuestos volátiles menos comunes. Un descubrimiento típico de la mosca es el vapor ligeramente alcohólico y fermentado que emana de un pino cuyas agujas exudan la savia que ha succionado una banda de áfidos. Las viejas gotas de resina se hallan en fermentación, cuyo olor atrae a la mosca. Un paseo, un minuto de succión, un paseo, una repetición. Tras unas pocas gotitas, la mosca se siente satisfecha. Los días de verano pasan sin ninguna actividad. Las moscas que no necesitan comer de inmediato ni están apareando "no van a parte alguna, ni hacen nada... La mayor parte del tiempo es de ocio sin rumbo", entre las moscas y entre la mayoría de los animales no sociales.

La mosca abastece su déficit energético con azúcares. "Tras el agua, la sacarosa es el compuesto más universalmente aceptado". El azúcar propio de la sangre de la mosca es un raro dímero de la glucosa, la trehalosa. Ocupa el centro de la explicación la elección y el rechazo de alimentos, el control de la ingestión y el gusto de las moscas. Los experimentadores han lanzado al vuelo moscas sujetas por un hilo para comprobar su hambre. Se han descubierto delicados circuitos nerviosos que pueden seccionarse para inducir a una mosca a comer sin mesura (en el libro aparecen moscas obesas) y han acoplado dos moscas, dorso con dorso, fundiendo sus torrentes sanguíneos para ver si una quedaba hambrienta después de haber sido alimentada su compañera (no quedaba). El cuerpo de una mosca adulta es extraño: no está proyectado de forma que crezca o se substituya alguna parte del mismo; en la vejez se torna simplemente ajada y harapienta, cual viejo gabán. No le crece pelo nuevo ni se restañan sus heridas; la mosca responde a la herida como lo hace un latón ante un soldador: mediante un coágulo en la lesión.

A las moscas se les puede condicionar. Margaret C. Nelson demostró que a una moscarda se le puede entrenar para que estire la trompa pidiendo un premio azucarado, dando a sus pies una señal de agua salada. En general, el aprendizaje es débil en las moscas; carecen de interacciones sociales, de nidos y de fuentes fijas de alimento. ¿Para qué aprender? Un pino es igual a otro; quizás, cuando el aprendizaje no tiene gran valor, el pequeño cerebro se ahorra para una computadora de fines específicos. El vuelo, la búsqueda, la elección y oportunidad de las comidas son una cantidad de necesidades intrínsecas para el cálculo. Las moscas no llegan jamás a acostumbrase a un saludo de la mano humana; siempre reaccionan volando, hasta que el hambre las deja inmovilizadas. Sin embargo, dejan de comer azúcar después de consumirlo en exceso. (Aún no es seguro si no se trata tan sólo de una muestra de fatiga del aparato sensorial). Este hermoso, costoso, compendiado y erudito volumen está escrito con ingenio y claridad, aunque es sumamente técnico en muchos capítulos. (Hay, por ejemplo, un atlas del cerebro de la moscarda, de

30 páginas.) El autor difícilmente habría podido comenzar su libro, ni este comentarista terminar su reseña, sin una cita de algún poeta metafísico (en este caso William Oldys): "Atareada, curiosa, sedienta mosca, /Bebe conmigo, y bebe como yo;..../ Sácale el jugo como puedas a la vida,/ La vida es breve y se desgasta.../ Sesenta veranos, cuando pasen, parecerán breves como uno solo".

LA FUNCION DE LA CIENCIA EN LA SOCIEDAD, Faustino Cordón, Editorial Cuadernos para el Diálogo, S.A., Madrid 1976. La sociología de la ciencia es un capítulo particular de la sociología del conocimiento, aquél que se ocupa de los principales logros del hombre que le permiten nuevas visiones del mundo y mejor uso del mismo. Pero, a propósito de ella, se está generalizando entre nosotros una opinión que es, cuando menos, muy discutible: la que atribuye al marxismo, concretamente a Gramsci, la paternidad de la sociología de la ciencia y de la función del intelectual en el seno de la sociedad. Tal vez sea uno de los progenitores de una corriente de interpretación marxista (desarrollada especialmente por Lukacs), que considera la ciencia como un factor superestructural más dependiente de la infraestructura económica. Corriente en la que se insertan historiadores tan eméritos como Benjamín Farrington (sobre la ciencia griega), Hessen, Crowther y Bernal (sobre la ciencia inglesa) y Dirk Struik (sobre la ciencia norteamericana).

Hay otros sectores que no son de tal parecer, sino que entienden el rol social de la ciencia y del intelectual, como acicate de renovación y perfeccionamiento del mundo entorno. Por ejemplo, Jaeger, Taton o los autores de la reciente obra española *Historia de la medicina.* Sobre el papel a representar por el intelectual en el seno de la sociedad y, en particular, frente al poder, Maravall ha publicado páginas muy documentadas.

Otros quieren convertir la ciencia en un huerto cerrado, sin influencias foráneas ni exteriorizaciones propias, a no ser por canales de rigurosa jerarquización. Son los devotos de la ciencia pura, los que viven enfervorizados en su arcadia de cristal, como era el caso de Gotinga antes de la Segunda Guerra o, en cierta medida, el MIT actual. A lo máximo, mantienen una postura moralizante y pacifista por bonhomía, sin relación con ningún sistema filosófico o político (Einstein, Pauling).

Por último, un cuarto grupo, de reciente creación, está constituido por quienes pretenden poner puertas al campo de la ciencia de forma muy sibilina (negando subvenciones a determinados tipos de investigación) o de forma más drástica y abierta (irrumpiendo por decreto ley en determinadas líneas de estudio: la creación *in vitro* de plásmidos, implantación de centrales nucleares), escudados siempre en una sonora "gestión pública de la ciencia y la tecnología".

El libro del bioquímico Faustino Cordón está a caballo entre el segundo y el tercer grupo. Pertenece al segundo por cuanto analiza los principales avances en el bienestar común que se ha conseguido a lo largo de la historia por obra y aplicación de la ciencia, y se numera en parte entre los que integran el tercer grupo por ser la suya una visión optimista y moralizante. "Como científicos herederos de la tradición experimental rechacemos decididamente la antinomia absurda de que la realidad es inteligible, racional, y, en cambio, el progreso humano de ella derivado sea inaprensible y no encauzable por acción humana en dirección conveniente."

Esta obra, además, tiene un contexto biobibliográfico que conviene señalar. Para Cordón, la evolución no se ciñe a la especie, sea vegetal o animal, sino que comprende ella, el medio y la interrelación de una con otra en todas sus manifestaciones. Una de ellas es la ciencia. La ciencia será evolucionista, como la sociedad en que se da: ciencia evolucionista que él resume así: "la ciencia, para salir de su asfixia por la maraña de datos inconexos que ella misma produce, tiene que procurar dos resultados apremiantes: (1) conseguir someter a conocimiento experimental a todos los procesos naturales que objetivamente sean susceptibles de ello, y (2) elevar el conjunto de los conocimientos experimentales a un nuevo tipo de relaciones, de problemas, que ciertamente escapan al horizonte conceptual de la ciencia experimental, pero cuya solución es sólo posible sobre la base brindada, precisamente, por la ciencia experimental misma". La incidencia social de esta concepción es puramente epistemológica: todos los hombres tienen derecho a llegar a la verdad y a contribuir a ella, sin arrogancias ni dogmatismos exclusivistas.

La preocupación científica en España es un fenómeno de la segunda mitad del siglo pasado. Antes hubo figuras aisladas que la promovieron, como el Padre Feijóo o las sociedades de Amigos del País. Se conoce un vigoroso impulso con la Residencia de Estudiantes y los primeros frutos del Instituto de Investigaciones Científicas, estos últimos, por desgracia, todavía muy menguados y sin continuidad. Falta, sin embargo, una opinión científica, como existe una opinión política.

TECNICOS, CIENTIFICOS Y CLASES SOCIALES, Daniel Lacalle, Guadarrama, Barcelona 1976. Según el aprecio que se profese a la ciencia en la sociedad así será el respeto público del hombre que se dedica a ella. Y, como una extensión, cabría decir lo propio del técnico. Los científicos y los técnicos constituían en todas partes, hasta ayer, la élite de las clases sociales.

Pero la generalización del bienestar y el ascenso de las clases medias a los sitios de formación y decisión ha dado al traste con ese hermetismo estructural y ha originado un nuevo fenómeno: la proletarización del investigador, del expositor y del técnico. Las así llamadas profesiones liberales se tornan serviles en el sentido de que se ejercen en un medio totalmente dependiente y marcado desde fuera y de antemano.

Los jóvenes graduados son los más afectados en esta revolución social: los nuevos ingenieros, médicos, abogados, biólogos, etcétera. Daniel Lacalle estudia dos casos pioneros y significativos: el caso Standard-ITT y el de los MIR; en el primero, son los ingenieros los interesados y, en el segundo, los médicos internos y residentes. Con la proletarización, llevan a la vida profesional una conciencia de clase en lucha, de sindicación, de reivindicación salarial y de cogestión en la empresa, cuyo alcance, todavía muy frenado por el sistema vigente en la mayoría de los países, es fácil prever: mayor fluidez y solidaridad con los demás asalariados, lucha por las mismas necesidades, por lograr las mismas mejoras y, al propio tiempo, un ataque sin piedad contra los monolitos sagrados (el colegio profesional, el sentido de élite, la ideología del éxito y el dominio).

Las crisis que se originan en la sociedad a partir de estas nuevas fuerzas están en el ambiente: paro, hipertrofia de técnicos, subempleo, ríos de dinero que se pierden en la formación de licenciados en especialidades totalmente innecesarias, desfasadas o saturadas. Y, sobre todo, falta un planteamiento serio del papel que debe cumplir el técnico y el científico en la sociedad del futuro, totalmente industrializada. Habrá que soñar nuevas autopías, nuevas repúblicas platonianas en donde cada pieza del engranaje social esté idealmente ocupada por el miembro idóneo.

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes;

EL ABORTO LEGAL

THE ABORTION EXPERIENCE: PSYCHOLOGICAL AND MEDICAL IMPACT. Dirigido por Howard J. Osofsky y Joy D. Osofsky. Harper and Row, Publishers, 1973.

LEGALIZED ABORTION AND THE PUBLIC HEALTH: REPORT OF A STUDY BY A COMMITTEE OF THE INSTITUTE OF MEDICINE, MAYO, 1975. Institute of Medicine. National Academy of Sciences, 1975.

ABORTION SURVEILLANCE: ANNUAL SUMMARY 1974. Center for Disease Control. Public Health Service, U.S. Department of Health, Education, and Welfare, 1976.

MORTALITY ASSOCIATED WITH THE CONTROL OF FERTILITY. Christopher Tietze, John Bongaarts y Bruce Schearer en *Family Planning Perspectives,* vol. 8, n.° 1, págs. 6-14; enero-febrero, 1976.

AGRICULTURA SIN LABRANZA

CONSERVATION TILLAGE: THE PROCEEDINGS OF A NATIONAL CONFERENCE SPONSORED BY THE SOIL CONSERVATION SOCIETY OF AMERICA IN COOPERATION WITH THE AMERICAN SOCIETY OF AGRONOMY, AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS, AND NATIONAL ASSOCIATION OF CONSERVATION DISTRICTS, marzo 28-30, 1973, en DES MOINES, IOWA. Soil Conservation Society of America, 1973.

INFLUENCE OF LONG TERM TILLAGE, CROP ROTATION, AND SOIL TYPE COMBINATIONS ON CORN YIELD. D. M. Van Doren, Jr., G. B. Triplett, Jr., y J. E. Henry en *Soil Science Society of America Journal,* vol. 40, n.° 1, págs. 100-105; enero-febrero, 1976.

LA MECANICA CUANTICA DE LOS AGUJEROS NEGROS

PARTICLE CREATION BY BLACK HOLES. S. W. Hawking en *Communications in Mathematical Physics,* vol. 43, n.° 3, págs. 199-220; 1975.

BLACK HOLES AND THERMODYNAMICS. S. W. Hawking en *Physical Review D,* vol. 13, n.° 2, págs. 191-197; enero 15, 1976.

EL LADO DE UNION DE LOS ANTICUERPOS

THREE-DIMENSIONAL STRUCTURE OF IMMUNOGLOBULINS. David R. Davies, Eduardo A. Padlan y David M. Segal en *Annual Review of Biochemistry,* vol. 44, págs. 639-667; 1975.

THREE-DIMENSIONAL STRUCTURE, FUNCTION AND GENETIC CONTROL OF IMMUNOGLOBULINS. Roberto J. Poljak en *Nature,* vol. 256, n.° 5516, págs. 373-376; julio 31, 1975.

ROTATIONAL ALLOMERISM AND DIVERGENT EVOLUTION OF DOMAINS IN IMMUNOGLOBULIN LIGHT CHAINS. A. B. Edmundson, K. R. Ely, E. E. Abola, M. Schiffer y N. Panagiotopoulos en *Biochemistry,* vol. 14, n.° 18, págs. 3953-3961; septiembre 9, 1975.

LA PERCEPCION DE OBJETOS EN MOVIMIENTO

MOVING VISUAL PHANTOMS: A NEW CONTOUR COMPLETION EFFECT. Paul Tynan y Robert Sekuler en *Science,* vol. 188, n.° 4191, págs. 951-952; mayo 30, 1975.

THE DETECTION OF FATIGUE DAMAGE BY EXOELECTRON EMISSION. William J. Baxter en *Journal of Applied Physics,* vol. 44, n.° 2, págs. 608-614; febrero, 1973.

NEUE ENTWICKLUNGSTENDENZEN IN DER ANWENDUNG DER EXOELEKTRONENEMISSION (FORTSCHRITTSBERICHT). W. Lausch en *Kernenergie,* vol. 17, n.° 5, págs. 129-136; mayo, 1974.

THE INDEPENDENCE OF CHANNELS IN HUMAN VISION SELECTIVE FOR DIRECTION OF MOVEMENT. Eugene Levinson y Robert Sekuler en *The Journal of Physiology,* vol. 250, n.° 2, págs. 347-366; septiembre, 1975.

EXOELECTRONES

A SURVEY OF EXO-ELECTRON EMISSION PHENOMENA. L. Grunberg en *British Journal of Applied Physics,* vol. 9, n.° 3, págs. 85-93; marzo, 1958.

REGISTRATION AND MAPPING OF PLASTICALLY DEFORMED METAL SURFACES BY MEANS OF PHOTOELECTRONS. C. C. Veerman en *Materials Science and Engineering,* vol. 4, n.° 6, págs. 329-342; septiembre, 1969.

FORMACION DE CRATERES EN EL SISTEMA SOLAR

MOONS AND PLANETS: AN INTRODUCTION TO PLANETARY SCIENCE. William K. Hartmann. Bogden & Quigley, Inc., Publishers, 1972.

PLANETARY GEOLOGY. Nicholas M. Short. Prentice-Hall, Inc., 1975.

MOON MORPHOLOGY: INTERPRETATIONS BASED ON LUNAR ORBITER PHOTOGRAPHY. Peter H. Schultz. University of Texas Press, 1976.

LA PRODUCTIVIDAD DE PRADOS MEDITERRANEOS

GRASSLAND ECOLOGY. C.R.W. Spedding Oxford. Clarendon Press, 1971.

PHOTOSYNTHESE ET PRODUCTION VEGETALE. C. Costes, Gauthier-Villars, C. Bordas, 1975.

FORAGE FERTILIZATION. D. A. Mays. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 1974.

CROP PROCESSES IN CONTROLLED ENVIRONMENTS. A. R. Rees, K. E. Cockshull, D. W. Hand y R. G. Hurd. Academic Press, London and New York, 1972.

CO_2 METABOLISM AND PLANT PRODUCTIVITY. R. H. Burris, y C. C. Black. University Park Press, Baltimore, London, Tokyo, 1975.

PHOTOSYNTHESIS AND PRODUCTIVITY IN DIFFERENT ENVIRONMENTS. J. P. Cooper. Cambridge University Press, Cambridge, London, New York, Melbourne, 1975.

JUEGOS MATEMATICOS

THE UNDECIDABILITY OF THE DOMINO PROBLEM. Robert Berger en *Memoirs of the American Mathematical Society,* n.° 66; 1966.

UNDECIDABILITY AND NONPERIODICITY FOR TILINGS OF THE PLANE. Raphael M. Robinson en *Inventiones Mathematicae,* vol. 12, fasc. 3, págs. 177-209; 1971.

THE ROLE OF AESTHETICS IN PURE AND APPLIED MATHEMATICAL RESEARCH. Roger Penrose en *Journal of the Institute of Mathematics and Its Applications,* vol. 10, n.° 7-8, págs. 266-271; julio-agosto, 1974.